

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46227 A2(51) International Patent Classification⁷: C07K 14/605, 14/765, C12N 15/62, A61K 38/38, 38/26, C07K 19/00, A61P 3/10

(21) International Application Number: PCT/US01/43165

(22) International Filing Date:
29 November 2001 (29.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/251,954 7 December 2000 (07.12.2000) US(71) Applicant (for all designated States except US): **ELI LILLY AND COMPANY** [US/US]; Lilly Corporate Center, Indianapolis, IN 46285 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GLAESNER, Wolfgang** [DE/US]; 7512 Fieldstone Court, Indianapolis, IN 46254 (US). **MICANOVIC, Radmilla** [US/US]; 7126 White Oak Trail, Indianapolis, IN 46236 (US). **TSCHANG, Sheng-Hung, Rainbow** [US/US]; 4963 Riley Mews, Carmel, IN 46033 (US).(74) Agents: **STEWART, Mark, J.** et al.; Eli Lilly and Company, Lilly Corporate Center, Drop Code 1104, Indianapolis, IN 46285 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, CZ (utility model), DE, DK, DK (utility model), DM, DZ, EC, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/46227 A2

(54) Title: GLP-1 FUSION PROTEINS

(57) Abstract: The present invention relates to glucagon-like-1 compounds fused to proteins that have the effect of extending the in vivo half-life of the peptides. These fusion proteins can be used to treat non-insulin dependent diabetes mellitus as well as a variety of other conditions.

- 1 -

GLP-1 FUSION PROTEINS

The present invention relates to glucagon-like peptides including analogs and derivatives thereof fused to proteins 5 that have the effect of extending the *in vivo* half-life of the peptides. These fusion proteins can be used to treat non-insulin dependent diabetes mellitus as well as a variety of other conditions.

10 Glucagon-Like Peptide 1 (GLP-1) is a 37 amino acid peptide that is secreted by the L-cells of the intestine in response to food ingestion. It has been found to stimulate insulin secretion (insulinotropic action), thereby causing glucose uptake by cells and decreased serum glucose levels 15 [see, e.g., Mojsov, S., (1992) *Int. J. Peptide Protein Research*, 40:333-343]. However, GLP-1 is poorly active. A subsequent endogenous cleavage between the 6th and 7th position produces a more potent biologically active GLP-1(7-37)OH peptide. Numerous GLP-1 analogs and derivatives are 20 known and are referred to herein as "GLP-1 compounds." These GLP-1 analogs include the Exendins which are peptides found in the venom of the GILA-monster. The Exendins have sequence homology to native GLP-1 and can bind the GLP-1 receptor and initiate the signal transduction cascade 25 responsible for the numerous activities that have been attributed to GLP-1(7-37)OH.

GLP-1 compounds have a variety of physiologically significant activities. For example, GLP-1 has been shown to stimulate insulin release, lower glucagon secretion, 30 inhibit gastric emptying, and enhance glucose utilization. [Nauck, M.A., et al. (1993) *Diabetologia* 36:741-744; Gutniak, M., et al. (1992) *New England J. of Med.* 326:1316-1322; Nauck, M.A., et al., (1993) *J. Clin. Invest.* 91:301-307].

- 2 -

GLP-1 shows the greatest promise as a treatment for non-insulin dependent diabetes mellitus (NIDDM). There are numerous oral drugs on the market to treat the insulin resistance associated with NIDDM. As the disease

5 progresses, however, patients must move to treatments that stimulate the release of insulin and eventually to treatments that involve injections of insulin. Current drugs which stimulate the release of insulin, however, can also cause hypoglycemia as can the actual administration of
10 insulin. GLP-1 activity, however, is controlled by blood glucose levels. When levels drop to a certain threshold level, GLP-1 is not active. Thus, there is no risk of hypoglycemia associated with treatment involving GLP-1.

15 However, the usefulness of therapy involving GLP-1 peptides has been limited by their fast clearance and short half-lives. For example, GLP-1(7-37) has a serum half-life of only 3 to 5 minutes. GLP-1(7-36) amide has a time action of about 50 minutes when administered subcutaneously. Even
20 analogs and derivatives that are resistant to endogenous protease cleavage, do not have half-lives long enough to avoid repeated administrations over a 24 hour period. Fast clearance of a therapeutic agent is inconvenient in cases where it is desired to maintain a high blood level of the agent over a prolonged period of time since repeated
25 administrations will then be necessary. Furthermore, a long-acting compound is particularly important for diabetic patients whose past treatment regimen has involved taking only oral medication. These patients often have an extremely difficult time transitioning to a regimen that
30 involves multiple injections of medication.

The present invention overcomes the problems associated with delivering a compound that has a short plasma half-life. The compounds of the present invention encompass GLP-1 compounds fused to another protein with a long circulating

- 3 -

half-life such as the Fc portion of an immunoglobulin or albumin.

Generally, small therapeutic peptides are difficult to manipulate because even slight changes in their structure 5 can affect stability and/or biological activity. This has been especially true for GLP-1 compounds currently in development. For example, GLP-1(7-37)OH has a tendency to undergo a conformational change from a primarily alpha helix structure to a primarily beta sheet structure. This beta 10 sheet form results in aggregated material that is thought to be inactive. It was, therefore, surprising that biologically active GLP-1 fusion proteins with increased half-lives could be developed. This was especially unexpected given the difficulty of working with GLP-1(7- 15 37)OH alone and the large size of the fusion partner relative to the small GLP-1 peptide attached.

Compounds of the present invention include heterologous fusion proteins comprising a first polypeptide with a N- 20 terminus and a C-terminus fused to a second polypeptide with a N-terminus and a C-terminus wherein the first polypeptide is a GLP-1 compound and the second polypeptide is selected from the group consisting of

- 25 a) human albumin;
- b) human albumin analogs; and
- c) fragments of human albumin,

and wherein the C-terminus of the first polypeptide is fused to the N-terminus of the second polypeptide.

Compounds of the present invention also include a 30 heterologous fusion protein comprising a first polypeptide with a N-terminus and a C-terminus fused to a second polypeptide with a N-terminus and a C-terminus wherein the first polypeptide is a GLP-1 compound and the second polypeptide is selected from the group consisting of

- 35 a) human albumin;

- 4 -

- b) human albumin analogs; and
- c) fragments of human albumin,

and wherein the C-terminus of the first polypeptide is fused to the N-terminus of the second polypeptide via a peptide linker. It is preferred that the peptide linker is selected from the group consisting of:

- a) a glycine rich peptide;
- b) a peptide having the sequence $[\text{Gly-Gly-Gly-Gly-Ser}]_n$ where n is 1, 2, 3, 4, 5 or 6; and
- c) a peptide having the sequence $[\text{Gly-Gly-Gly-Gly-Ser}]_3$.

Additional compounds of the present invention include a heterologous fusion protein comprising a first polypeptide with a N-terminus and a C-terminus fused to a second polypeptide with a N-terminus and a C-terminus wherein the first polypeptide is a GLP-1 compound and the second polypeptide is selected from the group consisting of

- a) the Fc portion of an immunoglobulin;
- b) an analog of the Fc portion of an immunoglobulin;

and

- c) fragments of the Fc portion of an immunoglobulin, and wherein the C-terminus of the first polypeptide is fused to the N-terminus of the second polypeptide. The GLP-1 compound may be fused to the second polypeptide via a peptide linker. It is preferable that the peptide linker is selected from the group consisting of:

- a) a glycine rich peptide;
- b) a peptide having the sequence $[\text{Gly-Gly-Gly-Gly-Ser}]_n$ where n is 1, 2, 3, 4, 5 or 6; and
- c) a peptide having the sequence $[\text{Gly-Gly-Gly-Gly-Ser}]_3$.

It is generally preferred that the GLP-1 compound that is part of the heterologous fusion protein have no more than 6 amino acids that are different from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4.

It is even more preferred that the GLP-1 compound have no

- 5 -

more than 5 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4. It is most preferred that the GLP-1 compound have no more than 4, 3, or 2 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4. Preferably, a GLP-1 compound that is part of the heterologous fusion protein has glycine or valine at position 8.

The present invention also includes polynucleotides encoding the heterologous fusion protein described herein, vectors comprising these polynucleotides and host cells transfected or transformed with the vectors described herein. Also included is a process for producing a heterologous fusion protein comprising the steps of transcribing and translating a polynucleotide described herein under conditions wherein the heterologous fusion protein is expressed in detectable amounts.

The present invention also encompasses a method for normalizing blood glucose levels in a mammal in need thereof comprising the administration of a therapeutically effective amount of a heterologous fusion protein described herein.

The invention is further illustrated with reference to the following drawings:

Figure 1: IgG1 Fc amino acid sequence encompassing the hinge region, CH2 and CH3 domains.

Figure 2: Human serum albumin amino acid sequence.

Figure 3: A. SDS-PAGE gel and immunoblot of same gel illustrating the molecular weight of IgG1-Fc and GLP-1-Fc fusion proteins (Lane 1, MW standards; Lane 2, Purified Fc; lane 3, Mock transfected media; Lane 4, Val⁸-GLP-1-Fc; Lane 5, Exendin-4-Fc) B. SDS-PAGE gel and immunoblot of same gel illustrating the molecular weight of human HSA and GLP-1-HSA fusion proteins (Lane 1, MW standards; Lane 2, Purified HSA; lane 3, Mock transfected media; Lane 4, Val⁸-GLP-1-HSA; Lane

- 6 -

5, Val⁸-GLP-1-[Gly-Gly-Gly-Gly-Ser]₃-HSA; Lane 6, Exendin-4-HSA; Lane 7, Exendin-4-[Gly-Gly-Gly-Gly-Ser]₃-HSA).

Figure 4: SDS-PAGE gel of purified Fc, albumin, and GLP-1 fusion proteins (Lane 1, MW standards; Lane 2, purified Fc; Lane 3, Val8-GLP-1-Fc; Lane 4, Exendin-4-Fc; Lane 5, MW standard; Lane 6, Val8-GLP-1-HSA; Lane 7, Exendin-4-HSA; Lane 8, Exendin-4-[Gly-Gly-Gly-Gly-Ser]₃-HSA).

Figure 5: Expression cloning vector containing the Fc regions illustrated in figure 1.

Figure 6: Expression cloning vector containing the albumin sequence illustrated in figure 2.

Figure 7: Expression cloning vector containing DNA encoding a 15 amino acid linker fused in frame and 5' of the albumin sequence illustrated in figure 2.

Figure 8: In vitro dose response activity of GLP-1 fusion proteins.

Figure 9: Pharmacokinetics of GLP-1 Fc and HSA fusion proteins.

Figure 10: Glucodynamic response to Exendin-Fc in two normal fasted dogs.

Figure 11: Insulinotropic response to Exendin-Fc in two normal fasted dogs.

Figure 12: DNA sequence encoding a human IgG1 Fc region.

Figure 13: DNA sequence encoding a human albumin protein.

The heterologous fusion proteins of the present invention comprise a GLP-1 compound fused to human albumin, a human albumin analog, a human albumin fragment, the Fc portion of an immunoglobulin, an analog of the Fc portion of an immunoglobulin, or a fragment of the Fc portion of an immunoglobulin. The C-terminus of the GLP-1 compound may be fused directly, or fused via a peptide linker, to the N-

- 7 -

terminus of an albumin or Fc protein. These heterologous fusion proteins are biologically active and have an increased half-life compared to native GLP-1.

It is preferred that the GLP-1 compounds that make up 5 part of the heterologous fusion protein encompass polypeptides having from about twenty-five to about thirty-nine naturally occurring or non-naturally occurring amino acids that have sufficient homology to native GLP-1(7-37)OH such that they exhibit insulinotropic activity by binding to 10 the GLP-1 receptor on β -cells in the pancreas. A GLP-1 compound typically comprises a polypeptide having the amino acid sequence of GLP-1(7-37)OH, an analog of GLP-1 (7-37)OH, a fragment of GLP-1(7-37)OH or a fragment of a GLP-1(7-37)OH analog. GLP-1(7-37)OH has the amino acid sequence of SEQ ID 15 NO: 1:

7	8	9	10	11	12	13	14	15	16	17
His-Ala-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val-Ser-										
18	19	20	21	22	23	24	25	26	27	28
Ser-Tyr-Leu-Glu-Gly-Gln-Ala-Ala-Lys-Glu-Phe-										
29	30	31	32	33	34	35	36	37		
Ile-Ala-Trp-Leu-Val-Lys-Gly-Arg-Gly										

(SEQ ID NO: 1)

By custom in the art, the amino terminus of GLP-1(7-25 37)OH has been assigned number residue 7 and the carboxy-terminus, number 37. The other amino acids in the polypeptide are numbered consecutively, as shown in SEQ ID NO: 1. For example, position 12 is phenylalanine and position 22 is glycine.

GLP-1 compounds also encompass "GLP-1 fragments." A 30 GLP-1 fragment is a polypeptide obtained after truncation of one or more amino acids from the N-terminus and/or C-terminus of GLP-1(7-37)OH or an analog or derivative thereof. The nomenclature used to describe GLP-1(7-37)OH is 35 also applicable to GLP-1 fragments. For example, GLP-1(9-

- 8 -

36)OH denotes a GLP-1 fragment obtained by truncating two amino acids from the *N*-terminus and one amino acid from the *C*-terminus. The amino acids in the fragment are denoted by the same number as the corresponding amino acid in GLP-1(7-37)OH. For example, the *N*-terminal glutamic acid in GLP-1(9-36)OH is at position 9; position 12 is occupied by phenylalanine; and position 22 is occupied by glycine, as in GLP-1(7-37)OH. For GLP-1(7-36)OH, the glycine at position 37 of GLP-1(7-37)OH is deleted.

10 GLP-1 compounds also include polypeptides in which one or more amino acids have been added to the *N*-terminus and/or *C*-terminus of GLP-1(7-37)OH, or fragments or analogs thereof. It is preferred that GLP-1 compounds of this type have up to about thirty-nine amino acids. The amino acids 15 in the "extended" GLP-1 compound are denoted by the same number as the corresponding amino acid in GLP-1(7-37)OH. For example, the *N*-terminus amino acid of a GLP-1 compound obtained by adding two amino acids to the *N*-terminal of GLP-1(7-37)OH is at position 5; and the *C*-terminus amino acid of 20 a GLP-1 compound obtained by adding one amino acid to the *C*-terminus of GLP-1(7-37)OH is at position 38. Thus, position 12 is occupied by phenylalanine and position 22 is occupied by glycine in both of these "extended" GLP-1 compounds, as in GLP-1(7-37)OH. Amino acids 1-6 of an extended GLP-1 25 compound are preferably the same as or a conservative substitution of the amino acid at the corresponding position of GLP-1(1-37)OH. Amino acids 38-45 of an extended GLP-1 compound are preferably the same as or a conservative substitution of the amino acid at the corresponding position 30 of glucagon or Exendin-4.

GLP-1 compounds of the present invention encompass "GLP-1 analogs." A GLP-1 analog has sufficient homology to GLP-1(7-37)OH or a fragment of GLP-1(7-37)OH such that the analog has insulinotropic activity. Preferably, a GLP-1 35 analog has the amino acid sequence of GLP-1(7-37)OH or a

- 9 -

fragment thereof, modified so that from one, two, three, four or five amino acids differ from the amino acid in the corresponding position of GLP-1(7-37)OH or a fragment of GLP-1(7-37)OH. In the nomenclature used herein to

5 designate GLP-1 compounds, the substituting amino acid and its position is indicated prior to the parent structure. For example, Glu²²-GLP-1(7-37)OH designates a GLP-1 compound in which the glycine normally found at position 22 of GLP-1(7-37)OH has been replaced with glutamic acid; Val⁸-Glu²²-
10 GLP-1(7-37)OH designates a GLP-1 compound in which alanine normally found at position 8 and glycine normally found at position 22 of GLP-1(7-37)OH have been replaced with valine and glutamic acid, respectively.

GLP-1 compounds of the present invention also include
15 "GLP-1 derivatives." A GLP-1 derivative is defined as a molecule having the amino acid sequence of GLP-1 or of a GLP-1 analog, but additionally having chemical modification of one or more of its amino acid side groups, α -carbon atoms, terminal amino group, or terminal carboxylic acid group. A chemical modification includes, but is not limited to, adding chemical moieties, creating new bonds, and removing chemical moieties. Modifications at amino acid side groups include, without limitation, acylation of lysine ϵ -amino groups, N-alkylation of arginine, histidine, or
20 lysine, alkylation of glutamic or aspartic carboxylic acid groups, and deamidation of glutamine or asparagine. Modifications of the terminal amino group include, without limitation, the des-amino, N-lower alkyl, N-di-lower alkyl, and N-acyl modifications. Modifications of the terminal
25 carboxy group include, without limitation, the amide, lower alkyl amide, dialkyl amide, and lower alkyl ester modifications. Lower alkyl is C₁-C₄ alkyl. Furthermore, one or more side groups, or terminal groups, may be
30 protected by protective groups known to the ordinarily-

- 10 -

skilled protein chemist. The α -carbon of an amino acid may be mono- or dimethylated.

Any GLP-1 compound can be part of the heterologous fusion proteins of the present invention as long as the GLP-1 compound itself is able to bind and induce signaling through the GLP-1 receptor. GLP-1 receptor binding and signal transduction can be assessed using *in vitro* assays such as those described in EP 619,322 and U.S. Patent No. 5,120,712, respectively.

10 Numerous active GLP-1 fragments, analogs and derivatives are known in the art and any of these analogs and derivatives can also be part of the heterologous fusion proteins of the present invention. Some examples of novel GLP-1 analogs as well as GLP-1 analogs and derivatives known 15 in the art are provided herein.

Some GLP-1 analogs and GLP-1 fragments known in the art include, for example, GLP-1(7-34) and GLP-1(7-35), GLP-1(7-36), Gln⁹-GLP-1(7-37), D-Gln⁹-GLP-1(7-37), Thr¹⁶-Lys¹⁸-GLP-1(7-37), and Lys¹⁸-GLP-1(7-37). GLP-1 analogs such as GLP-20 1(7-34) and GLP-1(7-35) are disclosed in U.S. Patent No. 5,118,666. Biologically processed forms of GLP-1 which have insulinotropic properties, such as GLP-1(7-36) are also known. Other known biologically active GLP-1 compounds are disclosed in U.S. Patent No 5,977,071 to Hoffmann, *et al.*, 25 U.S. Patent No. 5,545,618 to Buckley, *et al.*, and Adelhorst, *et al.*, *J. Biol. Chem.* 269:6275 (1994).

- 11 -

A preferred group of GLP-1 analogs is composed of GLP-1 analogs of formula I (SEQ ID NO: 2)

5 7 8 9 10 11 12 13 14 15 16 17
His-Xaa-Xaa-Gly-Xaa-Phe-Thr-Xaa-Asp-Xaa-Xaa-
18 19 20 21 22 23 24 25 26 27 28
Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Phe-
29 30 31 32 33 34 35 36 37 38 39
10 Ile-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-
40 41 42 43 44 45
Xaa-Xaa-Xaa-Xaa-Xaa-Xaa

Formula I (SEQ ID NO: 2)

15 wherein:

Xaa at position 8 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
Xaa at position 9 is Glu, Asp, or Lys;
Xaa at position 11 is Thr, Ala, Gly, Ser, Leu, Ile, Val,
20 Glu, Asp, or Lys;
Xaa at position 14 is Ser, Ala, Gly, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
Xaa at position 16 is Val, Ala, Gly, Ser, Thr, Leu, Ile,
Tyr, Glu, Asp, Trp, or Lys;
25 Xaa at position 17 is Ser, Ala, Gly, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
Xaa at position 18 is Ser, Ala, Gly, Thr, Leu, Ile, Val,
Glu, Asp, Trp, Tyr, or Lys;
Xaa at position 19 is Tyr, Phe, Trp, Glu, Asp, Gln, or Lys;
30 Xaa at position 20 is Leu, Ala, Gly, Ser, Thr, Ile, Val,
Glu, Asp, Met, Trp, Tyr, or Lys;
Xaa at position 21 is Glu, Asp, or Lys;
Xaa at position 22 is Gly, Ala, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
35 Xaa at position 23 is Gln, Asn, Arg, Glu, Asp, or Lys;

- 12 -

Xaa at position 24 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
Arg, Glu, Asp, or Lys;
Xaa at position 25 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
5 Xaa at position 26 is Lys, Arg, Gln, Glu, Asp, or His;
Xaa at position 27 is Leu, Glu, Asp, or Lys;
Xaa at position 30 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
Xaa at position 31 is Trp, Phe, Tyr, Glu, Asp, or Lys;
10 Xaa at position 32 is Leu, Gly, Ala, Ser, Thr, Ile, Val,
Glu, Asp, or Lys;
Xaa at position 33 is Val, Gly, Ala, Ser, Thr, Leu, Ile,
Glu, Asp, or Lys;
Xaa at position 34 is Asn, Lys, Arg, Glu, Asp, or His;
15 Xaa at position 35 is Gly, Ala, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
Xaa at position 36 is Gly, Arg, Lys, Glu, Asp, or His;
Xaa at position 37 is Pro, Gly, Ala, Ser, Thr, Leu, Ile,
Val, Glu, Asp, or Lys, or is deleted;
20 Xaa at position 38 is Ser, Arg, Lys, Glu, Asp, or His, or is
deleted;
Xaa at position 39 is Ser, Arg, Lys, Glu, Asp, or His, or is
deleted;
Xaa at position 40 is Gly, Asp, Glu, or Lys, or is deleted;
25 Xaa at position 41 is Ala, Phe, Trp, Tyr, Glu, Asp, or Lys,
or is deleted;
Xaa at position 42 is Ser, Pro, Lys, Glu, or Asp, or is
deleted;
Xaa at position 43 is Ser, Pro, Glu, Asp, or Lys, or is
30 deleted;
Xaa at position 44 is Gly, Pro, Glu, Asp, or Lys, or is
deleted;
and
Xaa at position 45 is Ala, Ser, Val, Glu, Asp, or Lys, or is
35 deleted;

- 13 -

provided that when the amino acid at position 37, 38, 39, 40, 41, 42, 43, or 44 is deleted, then each amino acid downstream of that amino acid is also deleted.

It is preferred that the GLP-1 compound of formula I 5 contain less than six amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH or Exendin-4. It is more preferred that less than five amino acids differ from the corresponding amino acid in GLP-1(7-37)OH or Exendin-4. It is even more preferred that less than four 10 amino acids differ from the corresponding amino acid in GLP-1(7-37)OH or Exendin-4.

GLP-1 compounds of the present invention include derivatives of formula I such as a C-1-6-ester, or amide, or C-1-6-alkylamide, or C-1-6-dialkylamide thereof. WO99/43706 15 describes derivatives of GLP-1 compounds of formula I and is incorporated by reference herein in its entirety. The compounds of formula I derivatized as described in WO99/43706 and underivatized are encompassed by the present invention.

20 Another preferred group of GLP-1 compounds is composed of GLP-1 analogs of formula II (SEQ ID NO: 3):

7 8 9 10 11 12 13 14 15 16 17
Xaa-Xaa-Xaa-Gly-Xaa-Xaa-Thr-Ser-Asp-Xaa-Ser-
25 18 19 20 21 22 23 24 25 26 27 28
Xaa-Xaa-Leu-Glu-Gly-Xaa-Xaa-Ala-Xaa-Xaa-Phe-
29 30 31 32 33 34 35 36 37
Ile-Xaa-Xaa-Leu-Xaa-Xaa-Xaa-R
Formula II (SEQ ID NO: 3)

30

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

- 14 -

Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;
Xaa at position 9 is: Thr, Ser, Arg, Lys, Trp, Phe, Tyr,
Glu, or His;
Xaa at position 11 is: Asp, Glu, Arg, Thr, Ala, Lys, or His;
5 Xaa at position 12 is: His, Trp, Phe, or Tyr;
Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp,
Val, Tyr, Glu, or Ala;
Xaa at position 18 is: His, Pro, Asp, Glu, Arg, Ser, Ala, or
Lys;
10 Xaa at position 19 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or
Cys;
Xaa at position 23 is: His, Asp, Lys, Glu, Gln, or Arg;
Xaa at position 24 is: Glu, Arg, Ala, or Lys;
Xaa at position 26 is: Trp, Tyr, Phe, Asp, Lys, Glu, or His;
15 Xaa at position 27 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, or
Lys;
Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;
Xaa at position 31 is: Asp, Glu, Ser, Thr, Arg, Trp, or Lys;
Xaa at position 33 is: Asp, Arg, Val, Lys, Ala, Gly, or Glu;
20 Xaa at position 34 is: Glu, Lys, or Asp;
Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe,
Asp, Gly, Pro, His, or Glu;
Xaa at position 36 is: Thr, Ser, Asp, Trp, Tyr, Phe, Arg,
Glu, or His;
25 R at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr,
Phe, His, Gly, Gly-Pro, or is deleted.

- 15 -

Another preferred group of GLP-1 compounds is composed of GLP-1 analogs of formula III (SEQ ID NO: 4):

7 8 9 10 11 12 13 14 15 16 17
5 Xaa-Xaa-Glu-Gly-Xaa-Xaa-Thr-Ser-Asp-Xaa-Ser-
18 19 20 21 22 23 24 25 26 27 28
Ser-Tyr-Leu-Glu-Xaa-Xaa-Xaa-Lys-Xaa-Phe-
29 30 31 32 33 34 35 36 37
Ile-Xaa-Trp-Leu-Xaa-Xaa-Xaa-R
10 formula III (SEQ ID NO: 4)

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine,

15 homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

Xaa at position 11 is: Asp, Glu, Arg, Thr, Ala, Lys, or His;

Xaa at position 12 is: His, Trp, Phe, or Tyr;

20 Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Glu, or Ala;

Xaa at position 22: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

25 Xaa at position 24 is: Glu, His, Ala, or Lys;

Xaa at position 25 is: Asp, Lys, Glu, or His;

Xaa at position 27 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, or Lys;

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

30 Xaa at position 33 is: Asp, Arg, Val, Lys, Ala, Gly, or Glu;

Xaa at position 34 is: Glu, Lys, or Asp;

Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe, Asp, Gly, Pro, His, or Glu;

Xaa at position 36 is: Arg, Glu, or His;

- 16 -

R at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

Another preferred group of GLP-1 compounds is composed
5 of GLP-1 analogs of formula IV (SEQ ID NO: 5):

7 8 9 10 11 12 13 14 15 16 17
Xaa-Xaa-Glu-Gly-Thr-Xaa-Thr-Ser-Asp-Xaa-Ser-
18 19 20 21 22 23 24 25 26 27 28
Ser-Tyr-Leu-Glu-Xaa-Xaa-Ala-Ala-Xaa-Glu-Phe-
10 29 30 31 32 33 34 35 36 37
Ile-Xaa-Trp-Leu-Val-Lys-Xaa-Arg-R
formula IV (SEQ ID NO: 5)

wherein:

15 Xaa at position 7 is: L-histidine, D-histidine, desamino-
histidine, 2-amino-histidine, β -hydroxy-histidine,
homohistidine, α -fluoromethyl-histidine or α -methyl-
histidine;
Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, Met, or
20 Thr;
Xaa at position 12 is: His, Trp, Phe, or Tyr;
Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp,
Val, Glu, or Ala;
Xaa at position 22 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or
25 Cys;
Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;
Xaa at position 26 is: Asp, Lys, Glu, or His;
Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;
Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe,
30 Asp, Gly, Pro, His, or Glu;
R at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr,
Phe, His, Gly, Gly-Pro, or is deleted.

- 17 -

Another preferred group of GLP-1 compounds is composed of GLP-1 analogs of formula V (SEQ ID NO: 6):

7 8 9 10 11 12 13 14 15 16 17
5 Xaa-Xaa-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val-Ser-
18 19 20 21 22 23 24 25 26 27 28
Ser-Tyr-Leu-Glu-Xaa-Xaa-Xaa-Ala-Lys-Glu-Phe-
29 30 31 32 33 34 35 36 37
Ile-Xaa-Trp-Leu-Val-Lys-Gly-Arg-R

10 formula V (SEQ ID NO: 6)

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

15 Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

Xaa at position 22 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

20 Xaa at position 24 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, or Lys;

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

R at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

25

Preferred GLP-1 compounds of formula I, II, III, IV, and V comprise GLP-1 analogs or fragments of GLP-1 analogs wherein the analogs or fragments contain an amino acid other than alanine at position 8 (position 8 analogs). It is preferable that these position 8 analogs contain one or more additional changes at positions 9, 11, 12, 16, 18, 22, 23, 24, 26, 27, 30, 31, 33, 34, 35, 36, and 37 compared to the corresponding amino acid of native GLP-1(7-37)OH. It is also preferable that these analogs have 6 or fewer changes compared to the corresponding amino acids in native GLP-1(7-

- 18 -

37)OH or GLP-1(7-36)OH. More preferred analogs have 5 or fewer changes compared to the corresponding amino acids in native GLP-1(7-37)OH or GLP-1(7-36)OH or have 4 or fewer changes compared to the corresponding amino acids in native 5 GLP-1(7-37)OH or GLP-1(7-36)OH. It is even more preferable that these analogs have 3 or fewer changes compared to the corresponding amino acids in native GLP-1(7-37)OH or GLP-1(7-36)OH. It is most preferable that these analogs have 2 or fewer changes compared to the corresponding amino acids 10 in native GLP-1(7-37)OH.

It has been found that the compounds of formula II, III, IV, and V have a reduced propensity to aggregate and generate insoluble forms. This is also important in the context of a fusion protein wherein the relatively small 15 GLP-1 peptide must maintain an active conformation despite being fused to a much larger protein. Preferred GLP-1 compounds of formula II, III, IV, and V encompassed by the fusion proteins of the present invention comprise GLP-1 analogs or fragments of GLP-1 analogs in which glycine at 20 position 22 and preferably alanine at position 8 have been replaced with another amino acid.

When position 22 is aspartic acid, glutamic acid, arginine or lysine, position 8 is preferably glycine, valine, leucine, isoleucine, serine, threonine or methionine 25 and more preferably valine or glycine. When position 22 is a sulfonic acid such as cysteic acid, position 8 is preferably glycine, valine, leucine, isoleucine, serine, threonine or methionine and more preferably valine or glycine.

30 Other preferred GLP-1 compounds include GLP-1 analogs of formula IV (SEQ ID NO:5) wherein the analogs have the sequence of GLP-1(7-37)OH except that the amino acid at position 8 is preferably glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more 35 preferably valine or glycine and position 30 is glutamic

- 19 -

acid, aspartic acid, serine, or histidine and more preferably glutamatic acid.

Other preferred GLP-1 compounds include GLP-1 analogs of formula IV (SEQ ID NO:5) wherein the analogs have the sequence of GLP-1(7-37)OH except that the amino acid at position 8 is preferably glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or glycine and position 37 is histidine, lysine, arginine, threonine, serine, glutamic acid, aspartic acid, tryptophan, tyrosine, phenylalanine and more preferably histidine.

Other preferred GLP-1 compounds include GLP-1 analogs of formula IV (SEQ ID NO:5) wherein the analogs have the sequence of GLP-1(7-37)OH except that the amino acid at position 8 is preferably glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or glycine and position 22 is glutamic acid, lysine, aspartic acid, or arginine and more preferably glutamic acid or lysine and position 23 is lysine, arginine, glutamic acid, aspartic acid, and histidine and more preferably lysine or glutamic acid.

Other preferred GLP-1 compounds include GLP-1 analogs of formula V (SEQ ID NO:6) wherein the analogs have the sequence of GLP-1(7-37)OH except that the amino acid at position 8 is preferably glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or glycine and position 22 is glutamic acid, lysine, aspartic acid, or arginine and more preferably glutamine acid or lysine and position 27 is alanine, lysine, arginine, tryptophan, tyrosine, phenylalanine, or histidine and more preferably alanine.

Other preferred GLP-1 compounds include GLP-1 analogs of formula II wherein the analogs have the sequence of GLP-

- 20 -

1(7-37)OH except that the amino acid at position 8 and one, two, or three amino acids selected from the group consisting of position 9, position 11, position 12, position 16, position 18, position 22, position 23, position 24, position 5 26, position 27, position 30, position 31, position 33, position 34, position 35, position 36, and position 37, differ from the amino acid at the corresponding position of native GLP-1(7-37)OH.

Other preferred GLP-1 compounds of formula II include:
10 Val⁸-GLP-1(7-37)OH, Gly⁸-GLP-1(7-37)OH, Glu²²-GLP-1(7-37)OH,
Asp²²-GLP-1(7-37)OH, Arg²²-GLP-1(7-37)OH, Lys²²-GLP-1(7-37)OH,
Cys²²-GLP-1(7-37)OH, Val⁸-Glu²²-GLP-1(7-37)OH, Val⁸-Asp²²-GLP-1(7-37)OH,
Val⁸-Arg²²-GLP-1(7-37)OH, Val⁸-Lys²²-GLP-1(7-37)OH, Val⁸-Cys²²-GLP-1(7-37)OH,
Gly⁸-Glu²²-GLP-1(7-37)OH, Gly⁸-Asp²²-GLP-1(7-37)OH,
Gly⁸-Lys²²-GLP-1(7-37)OH, Gly⁸-Cys²²-GLP-1(7-37)OH, Glu²²-GLP-1(7-36)OH,
Asp²²-GLP-1(7-36)OH, Arg²²-GLP-1(7-36)OH,
Lys²²-GLP-1(7-36)OH, Cys²²-GLP-1(7-36)OH, Val⁸-Glu²²-GLP-1(7-36)OH,
Val⁸-Asp²²-GLP-1(7-36)OH, Val⁸-Arg²²-GLP-1(7-36)OH,
Val⁸-Lys²²-GLP-1(7-36)OH, Val⁸-Cys²²-GLP-1(7-36)OH,
Gly⁸-Glu²²-GLP-1(7-36)OH, Gly⁸-Asp²²-GLP-1(7-36)OH, Gly⁸-Arg²²-GLP-1(7-36)OH,
Gly⁸-Lys²²-GLP-1(7-36)OH, Gly⁸-Cys²²-GLP-1(7-36)OH, Lys²³-GLP-1(7-37)OH,
Val⁸-Lys²³-GLP-1(7-37)OH, Gly⁸-Lys²³-GLP-1(7-37)OH, His²⁴-GLP-1(7-37)OH,
Val⁸-His²⁴-GLP-1(7-37)OH, Gly⁸-His²⁴-GLP-1(7-37)OH, Lys²⁴-GLP-1(7-37)OH,
Val⁸-Lys²⁴-GLP-1(7-37)OH, Gly⁸-Lys²³-GLP-1(7-37)OH, Glu³⁰-GLP-1(7-37)OH,
Val⁸-Glu³⁰-GLP-1(7-37)OH, Gly⁸-Glu³⁰-GLP-1(7-37)OH, Glu³⁰-GLP-1(7-37)OH,
Asp³⁰-GLP-1(7-37)OH, Val⁸-Asp³⁰-GLP-1(7-37)OH, Gly⁸-Asp³⁰-GLP-1(7-37)OH,
Gln³⁰-GLP-1(7-37)OH, Val⁸-Gln³⁰-GLP-1(7-37)OH, Gly⁸-Gln³⁰-GLP-1(7-37)OH,
Tyr³⁰-

- 21 -

GLP-1(7-37)OH, Val⁸-Tyr³⁰-GLP-1(7-37)OH, Gly⁸-Tyr³⁰-GLP-1(7-37)OH, Ser³⁰-GLP-1(7-37)OH, Val⁸-Ser³⁰-GLP-1(7-37)OH, Gly⁸-Ser³⁰-GLP-1(7-37)OH, His³⁰-GLP-1(7-37)OH, Val⁸-His³⁰-GLP-1(7-37)OH, Gly⁸-His³⁰-GLP-1(7-37)OH, Glu³⁴-GLP-1(7-37)OH, Val⁸-Glu³⁴-GLP-1(7-37)OH, Gly⁸-Glu³⁴-GLP-1(7-37)OH, Ala³⁴-GLP-1(7-37)OH, Val⁸-Ala³⁴-GLP-1(7-37)OH, Gly⁸-Ala³⁴-GLP-1(7-37)OH, Gly³⁴-GLP-1(7-37)OH, Val⁸-Gly³⁴-GLP-1(7-37)OH, Gly⁸-Gly³⁴-GLP-1(7-37)OH, Ala³⁵-GLP-1(7-37)OH, Val⁸-Ala³⁵-GLP-1(7-37)OH, Gly⁸-Ala³⁵-GLP-1(7-37)OH, Lys³⁵-GLP-1(7-37)OH, Val⁸-Lys³⁵-GLP-1(7-37)OH, Gly⁸-Lys³⁵-GLP-1(7-37)OH, His³⁵-GLP-1(7-37)OH Val⁸-His³⁵-GLP-1(7-37)OH, Gly⁸-His³⁵-GLP-1(7-37)OH, Pro³⁵-GLP-1(7-37)OH, Val⁸-Pro³⁵-GLP-1(7-37)OH, Gly⁸-Pro³⁵-GLP-1(7-37)OH, Glu³⁵-GLP-1(7-37)OH Val⁸-Glu³⁵-GLP-1(7-37)OH, Gly⁸-Glu³⁵-GLP-1(7-37)OH, Val⁸-Ala²⁷-GLP-1(7-37)OH, Val⁸-His³⁷-GLP-1(7-37)OH, Val⁸-Glu²²-Lys²³-GLP-1(7-37)OH, Val⁸-Glu²²-Ala²⁷-GLP-1(7-37)OH, Val⁸-Gly³⁴-Lys³⁵-GLP-1(7-37)OH, Val⁸-His³⁷-GLP-1(7-37)OH, Gly⁸-His³⁷-GLP-1(7-37)OH, Val⁸-Glu²²-Ala²⁷-GLP-1(7-37)OH, Gly⁸-Glu²²-Ala²⁷-GLP-1(7-37)OH, Val⁸-Lys²²-Glu²³-GLP-1(7-37)OH, and Gly⁸-Lys²²-Glu²³-GLP-1(7-37)OH.

Another preferred group of GLP-1 analogs and derivatives for use in the present invention is composed of molecules of formula VI (SEQ ID NO: 7)

25 **R₁-X-Glu-Gly¹⁰-Thr-Phe-Thr-Ser-Asp¹⁵-Val-Ser-Ser-Tyr-Leu²⁰-Y-Gly-Gln-Ala-Ala²⁵-Lys-**Z**-Phe-Ile-Ala³⁰-Trp-Leu-Val-Lys-Gly³⁵-Arg-**R₂****
 formula VI (SEQ ID NO:7)

- 22 -

wherein: R₁ is selected from the group consisting of L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, alpha-fluoromethyl-histidine, and alpha-methyl-histidine; X is selected from the group consisting of Ala, Gly, Val, Thr, Ile, and alpha-methyl-Ala; Y is selected from the group consisting of Glu, Gln, Ala, Thr, Ser, and Gly; Z is selected from the group consisting of Glu, Gln, Ala, Thr, Ser, and Gly; and R₂ is Gly-OH.

Another preferred group of GLP-1 compounds for use in the present invention is disclosed in WO 91/11457, and consists essentially of GLP-1(7-34), GLP-1(7-35), GLP-1(7-36), or GLP-1(7-37), or the amide form thereof, and pharmaceutically-acceptable salts thereof, having at least one modification selected from the group consisting of:

(a) substitution of glycine, serine, cysteine, threonine, asparagine, glutamine, tyrosine, alanine, valine, isoleucine, leucine, methionine, phenylalanine, arginine, or D-lysine for lysine at position 26 and/or position 34; or
substitution of glycine, serine, cysteine, threonine, asparagine, glutamine, tyrosine, alanine, valine, isoleucine, leucine, methionine, phenylalanine, lysine, or a D-arginine for arginine at position 36;

(b) substitution of an oxidation-resistant amino acid for tryptophan at position 31;

(c) substitution of at least one of: tyrosine for valine at position 16; lysine for serine at position 18; aspartic acid for glutamic acid at position 21; serine for glycine at position 22; arginine for glutamine at position 23; arginine for alanine at position 24; and glutamine for lysine at position 26; and

(d) substitution of at least one of: glycine, serine, or cysteine for alanine at position 8; aspartic acid, glycine, serine, cysteine, threonine, asparagine, glutamine, tyrosine, alanine, valine, isoleucine, leucine,

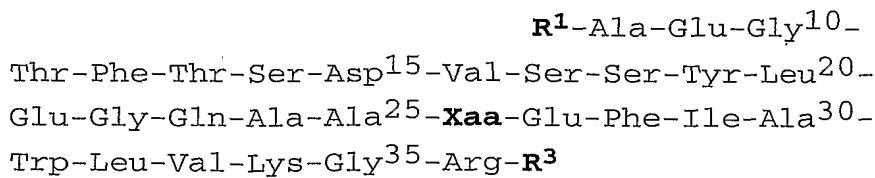
- 23 -

methionine, or phenylalanine for glutamic acid at position 9; serine, cysteine, threonine, asparagine, glutamine, tyrosine, alanine, valine, isoleucine, leucine, methionine, or phenylalanine for glycine at position 10; and glutamic acid for aspartic acid at position 15; and

(e) substitution of glycine, serine, cysteine, threonine, asparagine, glutamine, tyrosine, alanine, valine, isoleucine; leucine, methionine, or phenylalanine, or the D- or N-acylated or alkylated form of histidine for histidine at position 7; wherein, in the substitutions is (a), (b), (d), and (e), the substituted amino acids can optionally be in the D-form and the amino acids substituted at position 7 can optionally be in the N-acylated or N-alkylated form.

Because the enzyme, dipeptidyl-peptidase IV (DPP IV), may be responsible for the observed rapid *in vivo* inactivation of administered GLP-1, [see, e.g., Mentlein, R., et al., *Eur. J. Biochem.*, 214:829-835 (1993)], GLP-1 analogs and derivatives that are protected from the activity of DPP IV in the context of a fusion protein are preferred, and fusion proteins wherein the GLP-1 compound is Gly⁸-GLP-1(7-37)OH, Val⁸-GLP-1(7-37)OH, α -methyl-Ala⁸-GLP-1(7-37)OH, or Gly⁸-Gln²¹-GLP-1(7-37)OH are more preferred.

Another preferred group of GLP-1 compounds for use in the present invention consists of the compounds of formula VII (SEQ ID NO: 8) claimed in U.S. Patent No. 5,512,549, which is expressly incorporated herein by reference.



formula VII (SEQ ID NO: 8)

- 24 -

wherein R¹ is selected from the group consisting of 4-imidazopropionyl, 4-imidazoacetyl, or 4-imidazo- α,α dimethyl-acetyl; R² is selected from the group consisting of C₆-C₁₀ unbranched acyl, or is absent; R³ is selected from the group consisting of Gly-OH or NH₂; and, 5 Xaa is Lys or Arg.

More preferred compounds of formula IV for use in the present invention are those in which Xaa is Arg and R² is C₆-C₁₀ unbranched acyl. Even more preferred compounds of 10 formula IV for use in the present invention are those in which Xaa is Arg, R² is C₆-C₁₀ unbranched acyl, and R³ is Gly-OH. Other highly-preferred compounds of formula IV for use in the present invention are those in which Xaa is Arg, 15 R² is C₆-C₁₀ unbranched acyl, R³ is Gly-OH, and R¹ is 4-imidazopropionyl. An especially preferred compound of formula IV for use in the present invention is that in which Xaa is Arg, R² is C₈ unbranched acyl, R³ is Gly-OH, and R¹ is 4-imidazopropionyl.

Preferably, the GLP-1 compounds comprise GLP-1 analogs 20 wherein the backbone for such analogs or fragments contains an amino acid other than alanine at position 8 (position 8 analogs). The backbone may also include L-histidine, D-histidine, or modified forms of histidine such as desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, 25 homohistidine, α -fluoromethyl-histidine, or α -methyl-histidine at position 7. It is preferable that these position 8 analogs contain one or more additional changes at positions 12, 16, 18, 19, 20, 22, 25, 27, 30, 33, and 37 compared to the corresponding amino acid of native GLP-1(7-30 37)OH. It is more preferable that these position 8 analogs contain one or more additional changes at positions 16, 18,

- 25 -

22, 25 and 33 compared to the corresponding amino acid of native GLP-1(7-37)OH.

In a preferred embodiment, the GLP-1 analog is GLP-1(7-37)OH wherein the amino acid at position 12 is selected from 5 the group consisting of tryptophan or tyrosine. It is more preferred that in addition to the substitution at position 12, the amino acid at position 8 is substituted with glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or glycine. It is 10 even more preferred that in addition to the substitutions at position 12 and 8, the amino acid at position 22 is substituted with glutamic acid.

In another preferred embodiment, the GLP-1 analog is GLP-1(7-37)OH wherein the amino acid at position 16 is selected from the group consisting of tryptophan, 15 isoleucine, leucine, phenylalanine, or tyrosine. It is more preferred that in addition to the substitution at position 16, the amino acid at position 8 is substituted with glycine, valine, leucine, isoleucine, serine, threonine, or 20 methionine and more preferably valine or glycine. It is even more preferred that in addition to the substitutions at position 16 and 8, the amino acid at position 22 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 16 and 8, the 25 amino acid at position 30 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 16 and 8, the amino acid at position 37 is substituted with histidine.

In another preferred embodiment, the GLP-1 analog is 30 GLP-1(7-37)OH wherein the amino acid at position 18 is selected from the group consisting of tryptophan, tyrosine, phenylalanine, lysine, leucine, or isoleucine, preferably tryptophan, tyrosine, and isoleucine. It is more preferred

- 26 -

that in addition to the substitution at position 18, the amino acid at position 8 is substituted with glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or glycine. It is 5 even more preferred that in addition to the substitutions at position 18 and 8, the amino acid at position 22 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 18 and 8, the amino acid at position 30 is substituted with glutamic acid. 10 It is also preferred that in addition to the substitutions at positions 18 and 8, the amino acid at position 37 is substituted with histidine

In another preferred embodiment, the GLP-1 analog is GLP-1(7-37)OH wherein the amino acid at position 19 is 15 selected from the group consisting of tryptophan or phenylalanine, preferably tryptophan. It is more preferred that in addition to the substitution at position 19, the amino acid at position 8 is substituted with glycine, valine, leucine, isoleucine, serine, threonine, or 20 methionine and more preferably valine or glycine. It is even more preferred that in addition to the substitutions at position 19 and 8, the amino acid at position 22 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 19 and 8, the 25 amino acid at position 30 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 19 and 8, the amino acid at position 37 is substituted with histidine

In another preferred embodiment, the GLP-1 analog is 30 GLP-1(7-37)OH wherein the amino acid at position 20 is phenylalanine, tyrosine, or tryptophan. It is more preferred that in addition to the substitution at position 20, the amino acid at position 8 is substituted with

- 27 -

glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or glycine. It is even more preferred that in addition to the substitutions at position 20 and 8, the amino acid at position 22 is
5 substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 20 and 8, the amino acid at position 30 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 20 and 8, the amino acid at position 37 is
10 substituted with histidine

In another preferred embodiment, the GLP-1 analog is GLP-1(7-37)OH wherein the amino acid at position 25 is selected from the group consisting of valine, isoleucine, and leucine, preferably valine. It is more preferred that
15 in addition to the substitution at position 25, the amino acid at position 8 is substituted with glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or glycine. It is even more preferred that in addition to the substitutions at position
20 25 and 8, the amino acid at position 22 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 25 and 8, the amino acid at position 30 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions
25 25 and 8, the amino acid at position 37 is substituted with histidine.

In another preferred embodiment, the GLP-1 analog is GLP-1(7-37)OH wherein the amino acid at position 27 is selected from the group consisting of isoleucine or alanine.
30 It is more preferred that in addition to the substitution at position 27, the amino acid at position 8 is substituted with glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine or

- 28 -

glycine. It is even more preferred that in addition to the substitutions at position 27 and 8, the amino acid at position 22 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 5 27 and 8, the amino acid at position 30 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 27 and 8, the amino acid at position 37 is substituted with histidine

In another preferred embodiment, the GLP-1 analog is 10 GLP-1(7-37)OH wherein the amino acid at position 33 is isoleucine. It is more preferred that in addition to the substitution at position 33, the amino acid at position 8 is substituted with glycine, valine, leucine, isoleucine, serine, threonine, or methionine and more preferably valine 15 or glycine. It is even more preferred that in addition to the substitutions at position 33 and 8, the amino acid at position 22 is substituted with glutamic acid. It is also preferred that in addition to the substitutions at positions 33 and 8, the amino acid at position 30 is substituted with 20 glutamic acid. It is also preferred that in addition to the substitutions at positions 33 and 8, the amino acid at position 37 is substituted with histidine

The GLP-1 compounds have modifications at one or more of the following positions: 8, 12, 16, 18, 19, 20, 22, 25, 25 27, 30, 33, and 37. These GLP-1 compounds show increased potency compared with GLP-1(7-37)OH and comprise the amino acid sequence of formula IX (SEQ ID NO:12)

Xaa₇-Xaa₈-Glu-Gly-Thr-Xaa₁₂-Thr-Ser-Asp-Xaa₁₆-Ser-
Xaa₁₈-Xaa₁₉-Xaa₂₀-Glu-Xaa₂₂-Gln-Ala-Xaa₂₅-Lys-Xaa₂₇-
30 Phe-Ile-Xaa₃₀-Trp-Leu-Xaa₃₃-Lys-Gly-Arg-Xaa₃₇

Formula IX (SEQ ID NO: 12)

- 29 -

wherein:

Xaa₇ is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine,

5 homohistidine, α -fluoromethyl-histidine, or α -methyl-histidine;

Xaa₈ is: Ala, Gly, Val, Leu, Ile, Ser, or Thr;

Xaa₁₂ is: Phe, Trp, or Tyr;

Xaa₁₆ is: Val, Trp, Ile, Leu, Phe, or Tyr;

10 Xaa₁₈ is: Ser, Trp, Tyr, Phe, Lys, Ile, Leu, Val;

Xaa₁₉ is: Tyr, Trp, or Phe;

Xaa₂₀ is: Leu, Phe, Tyr, or Trp;

Xaa₂₂ is: Gly, Glu, Asp, or Lys;

Xaa₂₅ is: Ala, Val, Ile, or Leu;

15 Xaa₂₇ is: Glu, Ile, or Ala;

Xaa₃₀ is: Ala or Glu

Xaa₃₃ is: Val, or Ile; and

Xaa₃₇ is: Gly, His, NH₂, or is absent.

20 Some preferred GLP-1 compounds of formula IX include GLP-1(7-37)OH, GLP-1(7-36)-NH₂, Gly⁸-GLP-1(7-37)OH, Gly⁸-GLP-1(7-36)NH₂, Val⁸-GLP-1(7-37)OH, Val⁸-GLP-1(7-36)NH₂, Leu⁸-GLP-1(7-37)OH, Leu⁸-GLP-1(7-36)NH₂, Ile⁸-GLP-1(7-37)OH, Ile⁸-GLP-1(7-36)NH₂, Ser⁸-GLP-1(7-37)OH, Ser⁸-GLP-1(7-36)NH₂,
25 Thr⁸-GLP-1(7-37)OH, Thr⁸-GLP-1(7-36)NH₂, Val⁸-Tyr¹²-GLP-1(7-37)OH, Val⁸-Tyr¹²-GLP-1(7-36)NH₂, Val⁸-Tyr¹⁶-GLP-1(7-37)OH, Val⁸-Tyr¹⁶-GLP-1(7-36)NH₂, Val⁸-Glu²²-GLP-1(7-37)OH, Val⁸-Glu²²-GLP-1(7-36)NH₂, Gly⁸-Glu²²-GLP-1(7-37)OH, Gly⁸-Glu²²-GLP-1(7-36)NH₂, Val⁸-Asp²²-GLP-1(7-37)OH, Val⁸-Asp²²-GLP-1(7-36)NH₂, Gly⁸-Asp²²-GLP-1(7-37)OH, Gly⁸-Asp²²-GLP-1(7-36)NH₂, Val⁸-Lys²²-GLP-1(7-37)OH, Val⁸-Lys²²-GLP-1(7-36)NH₂, Gly⁸-Lys²²-GLP-1(7-37)OH, Gly⁸-Lys²²-GLP-1(7-36)NH₂, Leu⁸-Glu²²-GLP-1(7-37)OH, Leu⁸-Glu²²-GLP-1(7-36)NH₂, Ile⁸-Glu²²-GLP-1(7-37)OH, Ile⁸-Glu²²-GLP-1(7-36)NH₂.

- 30 -

37)OH, Ile⁸-Glu²²-GLP-1(7-36)NH₂, Leu⁸-Asp²²-GLP-1(7-37)OH, Leu⁸-Asp²²-GLP-1(7-36)NH₂, Ile⁸-Asp²²-GLP-1(7-37)OH, Ile⁸-Asp²²-GLP-1(7-36)NH₂, Leu⁸-Lys²²-GLP-1(7-37)OH, Leu⁸-Lys²²-GLP-1(7-36)NH₂, Ile⁸-Lys²²-GLP-1(7-37)OH, Ile⁸-Lys²²-GLP-1(7-36)NH₂, Ser⁸-Glu²²-GLP-1(7-37)OH, Ser⁸-Glu²²-GLP-1(7-36)NH₂, Thr⁸-Glu²²-GLP-1(7-37)OH, Thr⁸-Glu²²-GLP-1(7-36)NH₂, Ser⁸-Asp²²-GLP-1(7-37)OH, Ser⁸-Asp²²-GLP-1(7-36)NH₂, Thr⁸-Asp²²-GLP-1(7-37)OH, Thr⁸-Asp²²-GLP-1(7-36)NH₂, Ser⁸-Lys²²-GLP-1(7-37)OH, Ser⁸-Lys²²-GLP-1(7-36)NH₂, Thr⁸-Lys²²-GLP-1(7-37)OH, Ser⁸-Lys²²-GLP-1(7-36)NH₂, Thr⁸-Lys²²-GLP-1(7-37)OH, Thr⁸-Lys²²-GLP-1(7-36)NH₂, Glu²²-GLP-1(7-37)OH, Glu²²-GLP-1(7-36)NH₂, Asp²²-GLP-1(7-37)OH, Asp²²-GLP-1(7-36)NH₂, Lys²²-GLP-1(7-37)OH, Lys²²-GLP-1(7-36)NH₂, Val⁸-Ala²⁷-GLP-1(7-37)OH, Val⁸-Glu²²-Ala²⁷-GLP-1(7-37)OH, Val⁸-Glu³⁰-GLP-1(7-37)OH, Val⁸-Glu³⁰-GLP-1(7-36)NH₂, Gly⁸-Glu³⁰-GLP-1(7-37)OH, Gly⁸-Glu³⁰-GLP-1(7-36)NH₂, Leu⁸-Glu³⁰-GLP-1(7-37)OH, Leu⁸-Glu³⁰-GLP-1(7-36)NH₂, Ile⁸-Glu³⁰-GLP-1(7-37)OH, Ile⁸-Glu³⁰-GLP-1(7-36)NH₂, Ser⁸-Glu³⁰-GLP-1(7-37)OH, Ser⁸-Glu³⁰-GLP-1(7-36)NH₂, Thr⁸-Glu³⁰-GLP-1(7-37)OH, Thr⁸-Glu³⁰-GLP-1(7-36)NH₂, Val⁸-His³⁷-GLP-1(7-37)OH, Val⁸-His³⁷-GLP-1(7-36)NH₂, Gly⁸-His³⁷-GLP-1(7-37)OH, Gly⁸-His³⁷-GLP-1(7-36)NH₂, Leu⁸-His³⁷-GLP-1(7-37)OH, Leu⁸-His³⁷-GLP-1(7-36)NH₂, Ile⁸-His³⁷-GLP-1(7-37)OH, Ile⁸-His³⁷-GLP-1(7-36)NH₂, Ser⁸-His³⁷-GLP-1(7-37)OH, Ser⁸-His³⁷-GLP-1(7-36)NH₂, Thr⁸-His³⁷-GLP-1(7-37)OH, Thr⁸-His³⁷-GLP-1(7-36)NH₂.

Some preferred GLP-1 compounds of formula IX having multiple substitutions include GLP-1(7-37)OH wherein position 8 is valine or glycine, position 22 is glutamic acid, position 16 is tyrosine, leucine or tryptophan, position 18 is tyrosine, tryptophan, or isoleucine, position 25 is valine and position 33 is isoleucine. Other preferred GLP-1 compounds include the following: Val⁸-Tyr¹⁶-GLP-1(7-37)OH, Val⁸-Tyr¹²-Glu²²-GLP-1(7-37)OH, Val⁸-Tyr¹⁶-Phe¹⁹-GLP-1(7-37)OH, Val⁸-Tyr¹⁶-Glu²²-GLP-1(7-37)OH, Val⁸-Trp¹⁶-Glu²²-GLP-1(7-37)OH, Val⁸-Leu¹⁶-Glu²²-GLP-1(7-37)OH, Val⁸-Ile¹⁶-

- 31 -

Glu²²-GLP-1(7-37)OH, Val⁸-Phe¹⁶-Glu²²-GLP-1(7-37)OH; Val⁸-Trp¹⁸-Glu²²-GLP-1(7-37)OH, Val⁸-Tyr¹⁸-Glu²²-GLP-1(7-37)OH, Val⁸-Phe¹⁸-Glu²²-GLP-1(7-37)OH, and Val⁸-Ile¹⁸-Glu²²-GLP-1(7-37)OH.

5 The GLP-1 compounds of the present invention also encompass Exendin compounds. Exendin-3 and Exendin-4 are biologically active peptides first isolated from Helodermatidae lizard venoms and have been shown to bind the GLP-1 receptor and stimulate cAMP-dependent H⁺ production in 10 mammalian parietal cells. Exendin-3 and Exendin-4 are both 39 amino acid peptides which are approximately 53% homologous to GLP-1. They act as potent agonists of GLP-1 activity. Notably, an N-terminally truncated derivative of Exendin, known as Exendin(9-39 amino acids), is an inhibitor 15 of Exendin-3, Exendin-4 and GLP-1.

An Exendin compound typically comprises a polypeptide having the amino acid sequence of Exendin-3, Exendin-4, or an analog or fragment thereof. Exendin-3 and Exendin-4 are disclosed in U.S. Patent No. 5,424,286.

20

Exendin-3 has the amino acid sequence of SEQ ID NO: 9:

7	8	9	10	11	12	13	14	15	16	17	
His-Ser-Asp-Gly-Thr-Phe-Thr-Ser-Asp-Leu-Ser-											
18	19	20	21	22	23	24	25	26	27	28	
25	Lys-Gln-Met-Glu-Glu-Ala-Val-Arg-Leu-Phe-										
29	30	31	32	33	34	35	36	37	38	39	
Ile-Glu-Trp-Leu-Lys-Asn-Gly-Gly-Pro-Ser-Ser-											
40	41	42	43	44	45						
Gly-Ala-Pro-Pro-Pro-Ser											

30

(SEQ ID NO: 9)

- 32 -

Exendin-4 has the amino acid sequence of SEQ ID NO: 10:

7 8 9 10 11 12 13 14 15 16 17
His-Gly-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Leu-Ser-
5 18 19 20 21 22 23 24 25 26 27 28
Lys-Gln-Met-Glu-Glu-Glu-Ala-Val-Arg-Leu-Phe-
29 30 31 32 33 34 35 36 37 38 39
Ile-Glu-Trp-Leu-Lys-Asn-Gly-Gly-Pro-Ser-Ser-
40 41 42 43 44 45
10 Gly-Ala-Pro-Pro-Pro-Ser

(SEQ ID NO: 10)

GLP-1 compounds also include Exendin fragments which are polypeptides obtained after truncation of one or more 15 amino acids from the *N*-terminus and/or *C*-terminus of Exendin or an Exendin analog. Furthermore, GLP-1 compounds include Exendin polypeptides in which one or more amino acids have been added to the *N*-terminus and/or *C*-terminus of Exendin or fragments thereof. Exendin compounds of this type have up to 20 about forty-five amino acids.

GLP-1 compounds also include "Exendin analogs." An Exendin analog has sufficient homology to Exendin-4, Exendin-3, or a fragment thereof such that the analog has insulinotropic activity. The activity of Exendin fragments 25 and/or analogs can be assessed using *in vitro* assays such as those described in EP 619,322 and U.S. Patent No. 5,120,712.

Preferably, an Exendin analog has the amino acid sequence of Exendin-4 or a fragment thereof, modified so that from one, two, three, four or five amino acids differ 30 from the amino acid in corresponding position of Exendin-4 or the fragment of Exendin-4. In the nomenclature used herein to designate Exendin compounds, the substituting amino acid and its position is indicated prior to the parent

- 33 -

structure. For example, Val⁸-Exendin-4 designates an Exendin compound in which the glycine normally found at position 8 of Exendin-4 has been replaced with valine.

Another preferred group of GLP-1 compounds is composed 5 of GLP-1/Exendin-4 analogs of formula VIII (SEQ ID NO:11).

7 8 9 10 11 12 13 14 15 16 17
Xaa-Xaa-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Xaa-Ser-
18 19 20 21 22 23 24 25 26 27 28
10 Xaa-Xaa-Xaa-Glu-Xaa-Xaa-Ala-Xaa-Xaa-Xaa-Phe-
29 30 31 32 33 34 35 36 37
Ile-Xaa-Trp-Leu-Xaa-Xaa-Gly-Xaa-R
formula VIII (SEQ ID NO: 11)

15 wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

20 Xaa at position 8 is: Gly, Ala, or Val;

Xaa at position 16 is: Leu or Val;

Xaa at position 18 is Lys or Ser;

Xaa at position 19 is: Gln or Tyr;

Xaa at position 20 is: Met or Leu;

25 Xaa at position 22 is: Glu or Gln;

Xaa at position 23 is: Glu, or Gln;

Xaa at position 25 is: Val or Ala;

Xaa at position 26 is: Arg or Lys;

Xaa at position 27 is Leu or Glu;

30 Xaa at position 30 is: Glu or Ala;

Xaa at position 33 is: Val or Lys;

Xaa at position 34 is: Asn or Lys;

Xaa at position 36 is: Gly or Arg; and

- 34 -

R at position 37 is: Gly, Pro, Pro-Ser-Ser-Gly-Ala-Pro-Pro-Pro-Ser, or is absent. The activity of 18 different species that fall within this genus is provided in Table 6.

Further Exendin-analogs that are useful for the present invention are described in PCT patent publications WO 99/25728 (Beeley et al.), WO 99/25727 (Beeley et al.), WO 98/05351 (Young et al.), WO 99/40788 (Young et al.), WO 99/07404 (Beeley et al.), and WO 99/43708 (Knudsen et al.).

The GLP-1 fusion proteins of the present invention can comprise glycosylation sites. Glycosylation is a chemical modification wherein sugar moieties are added to the protein at specific sites. Glycosylation of proteins play a role in ensuring the correct charge, confirmation, and stability of maturing protein and can target the protein to the cell surface and eventual secretion of the protein. Most importantly, glycosylation effects the *in vivo* clearance rate for many proteins. Sugars can be O-linked or N-linked. Generally, O-linked sugars are added to the hydroxyl-group oxygen of serine and threonine, while N-linked sugars are added to the amide nitrogen of asparagine. The consensus site for N-glycosylation is Asn X1 X2 wherein X1 is any amino acid except Pro and X2 is Ser or Thr.

GLP-1 compounds are generally not glycosylated *in vivo*; however, interestingly the GLP-1 fusion proteins of the present invention that comprise a GLP-1 compound with a C terminal extension fused to an Fc sequence is glycosylated at the last serine in the C terminal extension (SSGAPPPS*) and at threonine at position 11 in the N terminal region of Fc (AEPKSCDKTHT*CPPC . . .).

Heterologous Fc fusion proteins:

The GLP-1 compounds described above can be fused directly or via a peptide linker to the Fc portion of an 5 immunoglobulin.

Immunoglobulins are molecules containing polypeptide chains held together by disulfide bonds, typically having two light chains and two heavy chains. In each chain, one domain (V) has a variable amino acid sequence depending on 10 the antibody specificity of the molecule. The other domains (C) have a rather constant sequence common to molecules of the same class.

As used herein, the Fc portion of an immunoglobulin has the meaning commonly given to the term in the field of 15 immunology. Specifically, this term refers to an antibody fragment which is obtained by removing the two antigen binding regions (the Fab fragments) from the antibody. One way to remove the Fab fragments is to digest the immunoglobulin with papain protease. Thus, the Fc portion 20 is formed from approximately equal sized fragments of the constant region from both heavy chains, which associate through non-covalent interactions and disulfide bonds. The Fc portion can include the hinge regions and extend through the CH₂ and CH₃ domains to the C-terminus of the antibody. 25 Representative hinge regions for human and mouse immunoglobulins can be found in Antibody Engineering, A Practical Guide, Borrebaeck, C.A.K., ed., W.H. Freeman and Co., 1992, the teachings of which are herein incorporated by reference. The Fc portion can further include one or more 30 glycosylation sites. The amino acid sequence of a representative Fc protein containing a hinge region, CH₂ and CH₃ domains, and one N-glycosylation site at position 82 is shown in Figure 1.

- 36 -

There are five types of human immunoglobulin Fc regions with different effect or and pharmacokinetic properties: IgG, IgA, IgM, IgD, and IgE. IgG is the most abundant immunoglobulin in serum. IgG also has the longest half-life in serum of any immunoglobulin (23 days). Unlike other immunoglobulins, IgG is efficiently recirculated following binding to an Fc receptor. There are four IgG subclasses G1, G2, G3, and G4, each of which have different effect or functions. G1, G2, and G3 can bind C1q and fix complement while G4 cannot. Even though G3 is able to bind C1q more efficiently than G1, G1 is more effective at mediating complement-directed cell lysis. G2 fixes complement very inefficiently. The C1q binding site in IgG is located at the carboxy terminal region of the CH2 domain.

All IgG subclasses are capable of binding to Fc receptors (CD16, CD32, CD64) with G1 and G3 being more effective than G2 and G4. The Fc receptor binding region of IgG is formed by residues located in both the hinge and the carboxy terminal regions of the CH2 domain.

IgA can exist both in a monomeric and dimeric form held together by a J-chain. IgA is the second most abundant Ig in serum, but it has a half-life of only 6 days. IgA has three effect or functions. It binds to an IgA specific receptor on macrophages and eosinophils, which drives phagocytosis and degranulation, respectively. It can also fix complement via an unknown alternative pathway.

IgM is expressed as either a pentamer or a hexamer, both of which are held together by a J-chain. IgM has a serum half-life of 5 days. It binds weakly to C1q via a binding site located in its CH3 domain. IgD has a half-life of 3 days in serum. It is unclear what effect or functions are attributable to this Ig. IgE is a monomeric Ig and has a serum half-life of 2.5 days. IgE binds to two Fc receptors which drives degranulation and results in the release of proinflammatory agents.

Depending on the desired *in vivo* effect, the heterologous fusion proteins of the present invention may contain any of the isotypes described above or may contain mutated Fc regions wherein the complement and/or Fc receptor binding functions have been altered. Thus, the heterologous fusion proteins of the present invention may contain the entire Fc portion of an immunoglobulin, fragments of the Fc portion of an immunoglobulin, or analogs thereof fused to a GLP-1 compound.

The fusion proteins of the present invention can consist of single chain proteins or as multi-chain polypeptides. Two or more Fc fusion proteins can be produced such that they interact through disulfide bonds that naturally form between Fc regions. These multimers can be homogeneous with respect to the GLP-1 compound or they may contain different GLP-1 compounds fused at the N-terminus of the Fc portion of the fusion protein.

Regardless of the final structure of the fusion protein, the Fc or Fc-like region must serve to prolong the *in vivo* plasma half-life of the GLP-1 compound fused at the N-terminus. Furthermore, the fused GLP-1 compound must retain some biological activity. An increase in half-life can be demonstrated using the method described in Example 7 wherein the half-life of the fusion protein is compared to the half-life of the GLP-1 compound alone. Biological activity can be determined by *in vitro* and *in vivo* methods known in the art. Representative biological assays are described in Examples 6, 8, and 9.

Since the Fc region of IgG produced by proteolysis has the same *in vivo* half-life as the intact IgG molecule and Fab fragments are rapidly degraded, it is believed that the relevant sequence for prolonging half-life reside in the CH2 and/or CH3 domains. Further, it has been shown in the literature that the catabolic rates of IgG variants that do not bind the high-affinity Fc receptor or C1q are

indistinguishable from the rate of clearance of the parent wild-type antibody, indicating that the catabolic site is distinct from the sites involved in Fc receptor or C1q binding. [Wawrzynczak et al., (1992) *Molecular Immunology* 29:221]. Site-directed mutagenesis studies using a murine IgG1 Fc region suggested that the site of the IgG1 Fc region that controls the catabolic rate is located at the CH2-CH3 domain interface.

Based on these studies, Fc regions can be modified at the catabolic site to optimize the half-life of the fusion proteins. It is preferable that the Fc region used for the heterologous fusion proteins of the present invention be derived from an IgG1 or an IgG4 Fc region. It is even more preferable that the Fc region be IgG4 or derived from IgG4. Preferably the IgG Fc region contains both the CH2 and CH3 regions including the hinge region.

Heterologous albumin fusion proteins:

The GLP-1 compounds described above can be fused directly or via a peptide linker to albumin or an analog, fragment, or derivative thereof.

Generally the albumin proteins making up part of the fusion proteins of the present invention can be derived from albumin cloned from any species. However, human albumin and fragments and analogs thereof are preferred to reduce the risk of the fusion protein being immunogenic in humans. Human serum albumin (HSA) consists of a single non-glycosylated polypeptide chain of 585 amino acids with a formula molecular weight of 66,500. The amino acid sequence of human HSA is shown in figure 2. [See Meloun, et al. (1975) FEBS Letters 58:136; Behrens, et al. (1975) Fed. Proc. 34:591; Lawn, et al. (1981) Nucleic Acids Research 9:6102-6114; Minghetti, et al. (1986) J. Biol. Chem. 261:6747]. A variety of polymorphic variants as well as analogs and fragments of albumin have been described. [See

- 39 -

Weitkamp, et al., (1973) Ann. Hum. Genet. 37:219]. For example, in EP 322,094, the inventors disclose various shorter forms of HSA. Some of these fragments include HSA(1-373), HSA(1-388), HSA(1-389), HSA(1-369), and HSA(1-419) and fragments between 1-369 and 1-419. EP 399,666 discloses albumin fragments that include HSA(1-177) and HSA(1-200) and fragments between HSA(1-177) and HSA(1-200).

It is understood that the heterologous fusion proteins of the present invention include GLP-1 compounds that are coupled to any albumin protein including fragments, analogs, and derivatives wherein such fusion protein is biologically active and has a longer plasma half-life than the GLP-1 compound alone. Thus, the albumin portion of the fusion protein need not necessarily have a plasma half-life equal to that of native human albumin. Fragments, analogs, and derivatives are known or can be generated that have longer half-lives or have half-lives intermediate to that of native human albumin and the GLP-1 compound of interest.

The heterologous fusion proteins of the present invention encompass proteins having conservative amino acid substitutions in the GLP-1 compound and/or the Fc or albumin portion of the fusion protein. A "conservative substitution" is the replacement of an amino acid with another amino acid that has the same net electronic charge and approximately the same size and shape. Amino acids with aliphatic or substituted aliphatic amino acid side chains have approximately the same size when the total number carbon and heteroatoms in their side chains differs by no more than about four. They have approximately the same shape when the number of branches in their side chains differs by no more than one. Amino acids with phenyl or substituted phenyl groups in their side chains are considered to have about the same size and shape. Except as otherwise specifically provided herein, conservative

- 40 -

substitutions are preferably made with naturally occurring amino acids.

However, the term "amino acid" is used herein in its broadest sense, and includes naturally occurring amino acids as well as non-naturally occurring amino acids, including amino acid analogs and derivatives. The latter includes molecules containing an amino acid moiety. One skilled in the art will recognize, in view of this broad definition, that reference herein to an amino acid includes, for example, naturally occurring proteogenic L-amino acids; D-amino acids; chemically modified amino acids such as amino acid analogs and derivatives; naturally occurring non-proteogenic amino acids such as norleucine, β -alanine, ornithine, GABA, etc.; and chemically synthesized compounds having properties known in the art to be characteristic of amino acids. As used herein, the term "proteogenic" indicates that the amino acid can be incorporated into a peptide, polypeptide, or protein in a cell through a metabolic pathway.

The incorporation of non-natural amino acids, including synthetic non-native amino acids, substituted amino acids, or one or more D-amino acids into the heterologous fusion proteins of the present invention can be advantageous in a number of different ways. D-amino acid-containing peptides, etc., exhibit increased stability *in vitro* or *in vivo* compared to L-amino acid-containing counterparts. Thus, the construction of peptides, etc., incorporating D-amino acids can be particularly useful when greater intracellular stability is desired or required. More specifically, D-peptides, etc., are resistant to endogenous peptidases and proteases, thereby providing improved bioavailability of the molecule, and prolonged lifetimes *in vivo* when such properties are desirable. Additionally, D-peptides, etc., cannot be processed efficiently for major histocompatibility complex class II-restricted presentation to T helper cells,

- 41 -

and are therefore, less likely to induce humoral immune responses in the whole organism.

In addition to structure/function analyses of the various polypeptides encompassed by the present invention,

5 there are numerous factors that can be considered when selecting amino acids for substitution. One factor that can be considered in making such changes is the hydropathic index of amino acids. The importance of the hydropathic amino acid index in conferring interactive biological 10 function on a protein has been discussed by Kyte and Doolittle (1982, *J. Mol. Biol.*, 157: 105-132). It is accepted that the relative hydropathic character of amino acids contributes to the secondary structure of the resultant protein. This, in turn, affects the interaction 15 of the protein with molecules such as enzymes, substrates, receptors, ligands, DNA, antibodies, antigens, etc. Based on its hydrophobicity and charge characteristics, each amino acid has been assigned a hydropathic index as follows:

isoleucine (+4.5); valine (+4.2); leucine (+3.8);
20 phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2);
glutamate/glutamine/aspartate/asparagine (-3.5); lysine 25 (-3.9); and arginine (-4.5).

As is known in the art, certain amino acids in a peptide, polypeptide, or protein can be substituted for other amino acids having a similar hydropathic index or score and produce a resultant peptide, etc., having similar 30 or even improved biological activity. In making such changes, it is preferable that amino acids having hydropathic indices within ± 2 are substituted for one another. More preferred substitutions are those wherein the amino acids have hydropathic indices within ± 1 . Most

- 42 -

preferred substitutions are those wherein the amino acids have hydropathic indices within ± 0.5 .

Like amino acids can also be substituted on the basis of hydrophilicity. U.S. Patent No. 4,554,101 discloses that 5 the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein. The following hydrophilicity values have been assigned to amino acids: arginine/lysine (+3.0); aspartate/glutamate (+3.0 ± 1); 10 serine (+0.3); asparagine/glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine/histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine/isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); and tryptophan (-3.4). Thus, one amino acid in a 15 peptide, polypeptide, or protein can be substituted by another amino acid having a similar hydrophilicity score and still produce a resultant peptide, etc., having similar biological activity, i.e., still retaining correct biological function. In making such changes, amino acids 20 having hydropathic indices within ± 2 are preferably substituted for one another, those within ± 1 are more preferred, and those within ± 0.5 are most preferred.

As outlined above, amino acid substitutions in the fusion proteins of the present invention can be based on the 25 relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, etc. Furthermore, substitutions can be made based on secondary structure propensity. For example, a helical amino acid can be 30 replaced with an amino acid that would preserve the helical structure. Exemplary substitutions that take various of the foregoing characteristics into consideration in order to produce conservative amino acid changes resulting in silent changes within the present peptides, etc., can be selected 35 from other members of the class to which the naturally

occurring amino acid belongs. Amino acids can be divided into the following four groups: (1) acidic amino acids; (2) basic amino acids; (3) neutral polar amino acids; and (4) neutral non-polar amino acids.

5

General methods for making the heterologous fusion proteins of the present invention.

Although the heterologous fusion proteins of the present invention can be made by a variety of different 10 methods, recombinant methods are preferred. For purposes of the present invention, as disclosed and claimed herein, the following general molecular biology terms and abbreviations are defined below. The terms and abbreviations used in this document have their normal meanings unless otherwise 15 designated. For example, "°C" refers to degrees Celsius; "mmol" refers to millimole or millimoles; "mg" refers to milligrams; "μg" refers to micrograms; "ml or mL" refers to milliliters; and "μl or μL" refers to microliters. Amino acids abbreviations are as set forth in 37 C.F.R. § 1.822 20 (b) (2) (1994).

"Base pair" or "bp" as used herein refers to DNA or RNA. The abbreviations A,C,G, and T correspond to the 5'-monophosphate forms of the deoxyribonucleosides (deoxy)adenosine, (deoxy)cytidine, (deoxy)guanosine, and 25 thymidine, respectively, when they occur in DNA molecules. The abbreviations U,C,G, and A correspond to the 5'-monophosphate forms of the ribonucleosides uridine, cytidine, guanosine, and adenosine, respectively when they occur in RNA molecules. In double stranded DNA, base pair 30 may refer to a partnership of A with T or C with G. In a DNA/RNA, heteroduplex base pair may refer to a partnership of A with U or C with G. (See the definition of "complementary", infra.)

"Digestion" or "Restriction" of DNA refers to the 35 catalytic cleavage of the DNA with a restriction enzyme that

- 44 -

acts only at certain sequences in the DNA ("sequence-specific endonucleases"). The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors, and other requirements were used as 5 would be known to one of ordinary skill in the art. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer or can be readily found in the literature.

"Ligation" refers to the process of forming 10 phosphodiester bonds between two double stranded nucleic acid fragments. Unless otherwise provided, ligation may be accomplished using known buffers and conditions with a DNA ligase, such as T4 DNA ligase.

"Plasmid" refers to an extrachromosomal (usually) self-replicating genetic element. Plasmids are generally 15 designated by a lower case "p" followed by letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available 20 plasmids in accordance with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Recombinant DNA cloning vector" as used herein refers 25 to any autonomously replicating agent, including, but not limited to, plasmids and phages, comprising a DNA molecule to which one or more additional DNA segments can or have been added.

"Recombinant DNA expression vector" as used herein 30 refers to any recombinant DNA cloning vector in which a promoter to control transcription of the inserted DNA has been incorporated.

"Transcription" refers to the process whereby 35 information contained in a nucleotide sequence of DNA is transferred to a complementary RNA sequence.

- 45 -

"Transfection" refers to the uptake of an expression vector by a host cell whether or not any coding sequences are, in fact, expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, 5 calcium phosphate co-precipitation, liposome transfection, and electroporation. Successful transfection is generally recognized when any indication of the operation of this vector occurs within the host cell.

"Transformation" refers to the introduction of DNA into 10 an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integration. Methods of transforming bacterial and eukaryotic hosts are well known in the art, many of which methods, such as nuclear injection, protoplast fusion or by calcium treatment 15 using calcium chloride are summarized in J. Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, (1989). Generally, when introducing DNA into Yeast the term transformation is used as opposed to the term transfection.

"Translation" as used herein refers to the process 20 whereby the genetic information of messenger RNA (mRNA) is used to specify and direct the synthesis of a polypeptide chain.

"Vector" refers to a nucleic acid compound used for the transfection and/or transformation of cells in gene 25 manipulation bearing polynucleotide sequences corresponding to appropriate protein molecules which, when combined with appropriate control sequences, confers specific properties on the host cell to be transfected and/or transformed. Plasmids, viruses, and bacteriophage are suitable vectors. 30 Artificial vectors are constructed by cutting and joining DNA molecules from different sources using restriction enzymes and ligases. The term "vector" as used herein includes Recombinant DNA cloning vectors and Recombinant DNA expression vectors.

- 46 -

"Complementary" or "Complementarity", as used herein, refers to pairs of bases (purines and pyrimidines) that associate through hydrogen bonding in a double stranded nucleic acid. The following base pairs are complementary: 5 guanine and cytosine; adenine and thymine; and adenine and uracil.

"Hybridization" as used herein refers to a process in which a strand of nucleic acid joins with a complementary strand through base pairing. The conditions employed in the 10 hybridization of two non-identical, but very similar, complementary nucleic acids varies with the degree of complementarity of the two strands and the length of the strands. Such techniques and conditions are well known to practitioners in this field.

15 "Isolated amino acid sequence" refers to any amino acid sequence, however, constructed or synthesized, which is locationally distinct from the naturally occurring sequence.

20 "Isolated DNA compound" refers to any DNA sequence, however constructed or synthesized, which is locationally distinct from its natural location in genomic DNA.

"Isolated nucleic acid compound" refers to any RNA or DNA sequence, however constructed or synthesized, which is locationally distinct from its natural location.

25 "Primer" refers to a nucleic acid fragment which functions as an initiating substrate for enzymatic or synthetic elongation.

"Promoter" refers to a DNA sequence which directs transcription of DNA to RNA.

30 "Probe" refers to a nucleic acid compound or a fragment, thereof, which hybridizes with another nucleic acid compound.

35 "Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In

- 47 -

general, longer probes require higher temperatures for proper annealing, while short probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to re-anneal when complementary strands are present in an 5 environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature that can be used. As a result, it follows that higher relative temperatures would tend to make the reactions more stringent, 10 while lower temperatures less so. For additional details and explanation of stringency of hybridization reactions, see Ausubel et al., *Current Protocols in Molecular Biology*, Wiley Interscience Publishers, 1995.

"Stringent conditions" or "high stringency conditions", 15 as defined herein, may be identified by those that (1) employ low ionic strength and high temperature for washing, for example, 15 mM sodium chloride/1.5 mM sodium citrate/0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for 20 example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride/75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5X SSC (750 mM sodium chloride, 75 mM sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5X Denhardt's 25 solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C with washes at 42°C in 0.2X SSC (30 mM sodium chloride/3 mM sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1X 30 SSC containing EDTA at 55°C.

"Moderately stringent conditions" may be identified as described by Sambrook et al. [*Molecular Cloning: A Laboratory Manual*, New York: Cold Spring Harbor Press, (1989)], and include the use of washing solution and hybridization 35 conditions (e.g., temperature, ionic strength, and %SDS) less

- 48 -

stringent than those described above. An example of moderately stringent conditions is overnight incubation at 37°C in a solution comprising: 20% formamide, 5X SSC (750 mM sodium chloride, 75 mM sodium citrate), 50 mM sodium phosphate 5 at pH 7.6, 5X Denhardt's solution, 10% dextran sulfate, and 20 mg/mL denatured sheared salmon sperm DNA, followed by washing the filters in 1X SSC at about 37-50°C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc., as necessary to accommodate factors such as probe length 10 and the like.

"PCR" refers to the widely-known polymerase chain reaction employing a thermally-stable DNA polymerase.

"Leader sequence" refers to a sequence of amino acids which can be enzymatically or chemically removed to produce 15 the desired polypeptide of interest.

"Secretion signal sequence" refers to a sequence of amino acids generally present at the N-terminal region of a larger polypeptide functioning to initiate association of that polypeptide with the cell membrane compartments like 20 endoplasmic reticulum and secretion of that polypeptide through the plasma membrane.

Construction of DNA encoding the heterologous fusion proteins of the present invention:

25 Wild-type albumin and Immunoglobulin proteins can be obtained from a variety of sources. For example, these proteins can be obtained from a cDNA library prepared from tissue or cells which express the mRNA of interest at a detectable level. Libraries can be screened with probes 30 designed using the published DNA or protein sequence for the particular protein of interest. For example, immunoglobulin light or heavy chain constant regions are described in Adams, et al. (1980) Biochemistry 19:2711-2719; Goughet, et al. (1980) Biochemistry 19:2702-2710; Dolby, et al. (1980) 35 Proc. Natl. Acad. Sci. USA 77:6027-6031; Rice et al. (1982)

- 49 -

Proc. Natl. Acad. Sci. USA 79:7862-7862; Falkner, et al. (1982) Nature 298:286-288; and Morrison, et al. (1984) Ann. Rev. Immunol. 2:239-256. Some references disclosing albumin protein and DNA sequences include Meloun, et al. (1975) FEBS Letters 58:136; Behrens, et al. (1975) Fed. Proc. 34:591; 5 Lawn, et al. (1981) Nucleic Acids Research 9:6102-6114; and Minghetti, et al. (1986) J. Biol. Chem. 261:6747

Screening a cDNA or genomic library with the selected probe may be conducted using standard procedures, such as 10 described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY (1989). An alternative means to isolate a gene encoding an albumin or immunoglobulin protein is to use PCR methodology [Sambrook et al., *supra*; Dieffenbach et al., *PCR Primer: A 15 Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY (1995)]. PCR primers can be designed based on published sequences.

Generally the full-length wild-type sequences cloned from a particular species can serve as a template to create 20 analogs, fragments, and derivatives that retain the ability to confer a longer plasma half-life on the GLP-1 compound that is part of the fusion protein. It is preferred that the Fc and albumin portions of the heterologous fusion proteins of the present invention be derived from the native 25 human sequence in order to reduce the risk of potential immunogenicity of the fusion protein in humans.

In particular, it is preferred that the immunoglobulin portion of a fusion protein encompassed by the present invention contain only an Fc fragment of the immunoglobulin. 30 Depending on whether particular effect or functions are desired and the structural characteristics of the fusion protein, an Fc fragment may contain the hinge region along with the CH2 and CH3 domains or some other combination thereof. These Fc fragments can be generated using PCR 35 techniques with primers designed to hybridize to sequences

- 50 -

corresponding to the desired ends of the fragment. Similarly, if fragments of albumin are desired, PCR primers can be designed which are complementary to internal albumin sequences. PCR primers can also be designed to create 5 restriction enzyme sites to facilitate cloning into expression vectors.

DNA encoding the GLP-1 compounds of the present invention can be made by a variety of different methods including cloning methods like those described above as well 10 as chemically synthesized DNA. Chemical synthesis may be attractive given the short length of the encoded peptide. The amino acid sequence for GLP-1 has been published as well as the sequence of the preproglucagon gene. [Lopez, et al. (1983) Proc. Natl. Acad. Sci., USA 80:5485-5489; Bell, et 15 al. (1983) Nature, 302:716-718; Heinrich, G., et al. (1984) Endocrinol, 115:2176-2181; Ghiglione, M., et al. 91984) Diabetologia 27:599-600]. Thus, primers can be designed to PCR native GLP-1 compounds and fragments thereof.

The gene encoding a fusion protein can then be 20 constructed by ligating DNA encoding a GLP-1 compound in-frame to DNA encoding an albumin or Fc protein. The gene encoding the GLP-1 compound and the gene encoding the albumin or Fc protein can also be joined in-frame via DNA encoding a linker peptide.

25 The *in vivo* function and stability of the heterologous fusion proteins of the present invention can be optimized by adding small peptide linkers to prevent potentially unwanted domain interactions. Although these linkers can potentially be any length and consist of any combination of amino acids, 30 it is preferred that the length be no longer than necessary to prevent unwanted domain interactions and/or optimize biological activity and/or stability. Generally, the linkers should not contain amino acids with extremely bulky side chains or amino acids likely to introduce significant 35 secondary structure. It is preferred that the linker be

- 51 -

serine-glycine rich and be less than 30 amino acids in length. It is more preferred that the linker be no more than 20 amino acids in length. It is even more preferred that the linker be no more than 15 amino acids in length. A 5 preferred linker contains repeats of the sequence Gly-Gly-Gly-Gly-Ser. It is preferred that there be between 2 and 6 repeats of this sequence. It is even more preferred that there be between 3 and 4 repeats of this sequence.

The DNA encoding wild-type GLP-1, albumin, and Fc 10 polypeptides and fragments thereof can be mutated either before ligation or in the context of a cDNA encoding an entire fusion protein. A variety of mutagenesis techniques are well known in the art. For example, a mutagenic PCR method utilizes strand overlap extension to create specific 15 base mutations for the purposes of changing a specific amino acid sequence in the corresponding protein. This PCR mutagenesis requires the use of four primers, two in the forward orientation (primers A and C) and two in the reverse orientation (primers B and D). A mutated gene is amplified 20 from the wild-type template in two different stages. The first reaction amplifies the gene in halves by performing an A to B reaction and a separate C to D reaction wherein the B and C primers target the area of the gene to be mutated. When aligning these primers with the target area, they 25 contain mismatches for the bases that are targeted to be changed. Once the A to B and C to D reactions are complete, the reaction products are isolated and mixed for use as the template for the A to D reaction. This reaction then yields the full, mutated product.

30 Once a gene encoding an entire fusion protein is produced it can be cloned into an appropriate expression vector. Specific strategies that can be employed to make the GLP-1 fusion proteins of the present invention are described in example 1.

General methods to recombinantly express the heterologous fusion proteins of the present invention:

Host cells are transfected or transformed with expression or cloning vectors described herein for heterologous fusion 5 protein production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without 10 undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in *Mammalian Cell Biotechnology: A Practical Approach*, M. Butler, ed. (IRL Press, 1991) and Sambrook, et al., *supra*. Methods of transfection are known to 15 the ordinarily skilled artisan, for example, CaPO_4 and electroporation. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of van Solingen et al., *J Bact.* 20 130(2): 946-7 (1977) and Hsiao et al., *Proc. Natl. Acad. Sci. USA* 76(8): 3829-33 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, e.g., polybrene or polyomithine, may also be 25 used. For various techniques for transforming mammalian cells, see Keown, et al., *Methods in Enzymology* 185: 527-37 (1990) and Mansour, et al., *Nature* 336(6197): 348-52 (1988).

Suitable host cells for cloning or expressing the nucleic acid (e.g., DNA) in the vectors herein include prokaryote, 30 yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 35 3 1.446); *E. coli* X1 776 (ATCC 3 1.537); *E. coli* strain W3 110

- 53 -

(ATCC 27.325) and K5 772 (ATCC 53.635). Other suitable prokaryotic host cells include Enterobacteriaceae such as Escherichia, e.g., *E. coli*, *Enterobacter*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, e.g., *Salmonella typhimurium*, *Serratia*, 5 e.g., *Serratia marcescans*, and *Shigeila*, as well as *Bacilli* such as *B. subtilis* and *B. licheniformis* (e.g., *B. licheniformis* 41 P disclosed in DD266,710, published 12 April 1989), *Pseudomonas* such as *P. aeruginosa*, and *Streptomyces*. These examples are illustrative rather than 10 limiting. Strain W3110 is one particularly preferred host or parent host because it is a common host strain for recombinant DNA product fermentations. Preferably, the host cell secretes minimal amounts of proteolytic enzymes. For example, strain W3 110 may be modified to effect a genetic mutation in the 15 genes encoding proteins endogenous to the host, with examples of such hosts including *E. coli* W3110 strain 1A2, which has the complete genotype *ronA*; *E. coli* W3 110 strain 9E4, which has the complete genotype *ton4 ptr3*; *E. coli* W3110 strain 27C7 (ATCC 55,244), which has the complete genotype *tonA, ptr3 phoA* 20 *E15 (argF-lac) I69 degP ompT /can'*; *E. coli* W3110 strain 40B4, which is strain 37D6 with a non-kanamycin resistant *degP* deletion mutation; and an *E. coli* strain having mutant 25 periplasmic protease disclosed in U.S. Patent No. 4,946,783 issued 7 August 1990. Alternatively, *in vivo* methods of cloning, e.g., PCR or other nucleic acid polymerase reactions, are suitable.

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for fusion protein vectors. *Saccharomyces cerevisiae* is 30 a commonly used lower eukaryotic host microorganism. Others include *Schizosaccharomyces pombe* [Beach and Nurse, *Nature* 290: 140-3 (1981); EP 139,383 published 2 May 1995]; *Muyveromyces* hosts [U.S. Patent No. 4,943,529; Fleer, et al., *Bio/Technology* 9(10): 968-75 (1991)] such as, e.g., *K lactis* 35 (MW98-8C, CBS683, CBS4574) [de Louvencourt et al., *J.*

- 54 -

Bacteriol. 154(2): 737-42 (1983)]; *K. fiagilis* (ATCC 12,424), *K. bulgaricus* (ATCC 16,045), *K. wickeramii* (ATCC 24,178), *K. waltii* (ATCC 56,500), *K. drosophilicarum* (ATCC 36,906) [Van den Berg et al., *Bio/Technology* 8(2): 135-9 (1990)]; *K.* 5 *thermotoierans*, and *K. marxianus*; *yarrowia* (EP 402,226); *Pichia pastoris* (EP 183,070) [Sreekrishna et al., *J. Basic Microbiol.* 28(4): 265-78 (1988)]; *Candid*; *Trichoderma reesiae* (EP 244,234); *Neurospora crassa* [Case, et al., *Proc. Natl. Acad. Sci. USA* 76(10): 5259-63 (1979)]; *Schwanniomyces* such as 10 *Schwanniomyces occidentalis* (EP 394,538 published 31 October 1990); and filamentous fungi such as, e.g., *Neurospora*, *Penicillium*, *Tolypocladium* (WO 91/00357 published 10 January 1991), and *Aspergillus* hosts such as *A. nidulans* [Ballance et al., *Biochem. Biophys. Res. Comm.* 112(1): 284-9 (1983)]; 15 *Tilburn*, et al., *Gene* 26(2-3): 205-21 (1983); *Yelton*, et al., *Proc. Natl. Acad. Sci. USA* 81(5): 1470-4 (1984)] and *A. niger* [Kelly and Hynes, *EMBO J.* 4(2): 475-9 (1985)]. Methylotrophic yeasts are selected from the genera consisting of *Hansenula*, *Candida*, *Kloeckera*, *Pichia*, *Saccharomyces*, *Torulopsis*, and 20 *Rhodotorula*. A list of specific species that are exemplary of this class of yeast may be found in C. Antony, *The Biochemistry of Methylotrophs* 269 (1982).

Suitable host cells for the expression of the fusion proteins of the present invention are derived from 25 multicellular organisms. Examples of invertebrate cells include insect cells such as *Drosophila S2* and *Spodoptera Sp*, *Spodoptera high5* as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney 30 CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line [293 or 293 cells subcloned for growth in suspension culture, Graham, et al., *J. Gen Virol.*, 36(1): 59-74 (1977)]; Chinese hamster ovary cells/-DHFR [CHO, Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77(7): 4216-20 35 (1980)]; mouse sertoli cells [TM4, Mather, *Biol. Reprod.*

- 55 -

23(1):243-52 (1980)]; human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

5 The fusion proteins of the present invention may be recombinantly produced directly, or as a protein having a signal sequence or other additional sequences which create a specific cleavage site at the N-terminus of the mature fusion protein. In general, the signal sequence may be a component of the vector, or it may be a part of the fusion protein-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders.

10 For yeast secretion the signal sequence may be, e.g., the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* cc-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179), or

15 the signal described in WO 90/13646. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species as well as viral secretory leaders.

20

25 Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2u plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

30

35 Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to

- 56 -

antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement autotrophic deficiencies, or (c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase
5 for Bacilli.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the fusion protein-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell
10 when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described [Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77(7): 4216-20 (1980)]. A suitable selection gene for use in yeast is the trpl gene present in the yeast plasmid YRp7 [Stinchcomb, et
15 al., *Nature* 282(5734): 39-43 (1979); Kingsman, et al., *Gene* 7(2): 141-52 (1979); Tschumper, et al., *Gene* 10(2): 157-66 (1980)]. The trpl gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEPC1 [Jones,
20 *Genetics* 85: 23-33 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the fusion protein-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters
25 suitable for use with prokaryotic hosts include the β -lactamase and lactose promoter systems [Chang, et al., *Nature* 275(5681): 617-24 (1978); Goeddel, et al., *Nature* 281(5732): 544-8 (1979)], alkaline phosphatase, a tryptophan (up) promoter system [Goeddel, *Nucleic Acids Res.* 8(18): 4057-74 (1980); EP 36,776 published 30 September 1981], and hybrid
30 promoters such as the tat promoter [deBoer, et al., *Proc. Natl. Acad. Sci. USA* 80(1): 21-5 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the fusion
35 protein.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman, et al., *J. Biol. Chem.* 255(24): 12073-80 (1980)] or other glycolytic enzymes [Hess et al., *J. Adv. Enzyme Reg.* 7: 149 (1968); Holland, *Biochemistry* 17(23): 4900-7 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

Suitable vectors and promoters for use in yeast expression are further described in EP 73,657. Transcription of fusion protein-encoding mRNA from vectors in mammalian host cells may be controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus, adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-

B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a polynucleotide encoding a fusion protein by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin,

- 58 -

elastase, albumin, α -ketoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early 5 promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the fusion protein coding sequence but is preferably located at a site 5' from the promoter.

10 Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available 15 from the 5' and occasionally 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding the fusion protein.

20 Various forms of a fusion protein may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (e.g., Triton-X 100) or by enzymatic cleavage. Cells employed in expression of a fusion protein 25 can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

30 Purification of the heterologous fusion proteins of the present invention:

Once the heterologous fusion proteins of the present invention are expressed in the appropriate host cell, the analogs can be isolated and purified. The following procedures are exemplary of suitable purification procedures: 35 fractionation on carboxymethyl cellulose; gel filtration such

as Sephadex G-75; anion exchange resin such as DEAE or Mono-Q; cation exchange such as CM or Mono-S; protein A sepharose to remove contaminants such as IgG; metal chelating columns to bind epitope-tagged forms of the polypeptide; reversed-phase 5 HPLC; chromatofocusing; silica gel; ethanol precipitation; and ammonium sulfate precipitation.

Various methods of protein purification may be employed and such methods are known in the art and described, for example, in Deutscher, *Methods in Enzymology* 182: 83-9 (1990) 10 and Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag, NY (1982). The purification step(s) selected will depend on the nature of the production process used and the particular fusion protein produced. For example, fusion 15 proteins comprising an Fc fragment can be effectively purified using a Protein A or Protein G affinity matrix. Low or high pH buffers can be used to elute the fusion protein from the affinity matrix. Mild elution conditions will aid in preventing irreversible denaturation of the fusion protein. 20 Imidazole-containing buffers can also be used. Example 3 describes some successful purification protocols for the fusion proteins of the present invention.

Characterization of the heterologous fusion proteins of the present invention:

25 Numerous methods exist to characterize the fusion proteins of the present invention. Some of these methods include: SDS-PAGE coupled with protein staining methods or immunoblotting using anti-IgG or anti-HSA antibodies. Other methods include matrix assisted laser desorption/ionization- 30 mass spectrometry (MALDI-MS), liquid chromatography/mass spectrometry, isoelectric focusing, analytical anion exchange, chromatofocusing, and circular dichroism to name a few. A representative number of heterologous fusion proteins were characterized using SDS-PAGE coupled with immunoblotting as

- 60 -

well as mass spectrometry (See examples 4 and 5 and figures 3 and 4).

For example table 3 (see example 5) illustrates the calculated molecular mass for a representative number of 5 fusion proteins as well as the mass as determined by mass spectrometry. In addition, Figures 3 and 4 illustrate molecular weights of a representative number of fusion proteins as determined by SDS PAGE. All heterologous fusion proteins tested were expressed and secreted transiently. In 10 addition, the Ig κ signal sequence was cleaved to yield proteins with the correct N-terminus.

Further, table 3 illustrates that in some instances the mass determined by mass spectrometry is greater than expected. This is the result of glycosylation of the Fc portion and the 15 C terminal extension. Enzymatic digestion of the fusion proteins followed by reversed-phase HPLC and mass spectrometry can identify peptide fractions that contain sugar moieties. These fractions can then be N-terminal amino acid sequenced to identify the potential glycosylation site. For example, 20 characterization of Exendin-4-Fc (SEQ ID NO: 29) shows that the serine at position 39 and threonine at position 50 are O-linked glycosylated and the asparagine at position 122 is N-linked glycosylated.

A representative number of GLP-1 fusion proteins were 25 also tested for activity. Numerous methods exist to detect GLP-1 activity *in vitro* and *in vivo* (see examples 6, 7, 8, and 9). Table 4 (example 6) illustrates GLP-1 receptor activity associated with several GLP-1 fusions. The numbers are relative to the activity associated with Val⁸-GLP-1(7-37)OH. 30 All fusion proteins tested had GLP-1 receptor activity. A low level of *in vitro* activity is not necessarily indicative of a weak effect *in vivo*. Because of the substantial increase in the half-life of these fusion proteins, weak *in vitro* activity is not generally a predictor of weak *in vivo* activity. Figure 35 7 and example 7 illustrate the prolonged half-life associated

- 61 -

with the fusion proteins of the present invention. For example, Val⁸-GLP-1-Fc had a half-life of approximately 45 hours in monkeys, Val⁸-GLP-1-HSA had a half-life of about 87 hours in monkeys, Gly⁸-Glu²²-GLP-1-CEx-Linker-IgG1 had a half-life after IV administration of approximately 55 hours in dogs, and Gly⁸-Glu²²-GLP-1-CEx-Linker-IgG1 had a half-life after SC administration of approximately 38 hours in dogs.

Compositions of the invention:

Physical stability is also an essential feature for therapeutic protein formulations. GLP-1 compounds have been especially difficult to manufacture and formulate due to structural changes that occur during processing. For example, some GLP-1 compounds have a general tendency to aggregate. In addition, it has been shown that some GLP-1 compounds convert from a soluble and active α -helix form to an insoluble and potentially inactive β -sheet form. The fusion of GLP-1 compounds to large proteins such as the Fc region of an IgG or albumin not only acts to increase the half-life of the GLP-1 compound, but also contributes to the physical and conformational stability of the GLP-1 compound. For example, Val⁸-GLP-1-Linker-HSA in PBS is stable at 37°C out to about 30 days.

The heterologous fusion proteins of the present invention may be formulated with one or more excipients. The active fusion proteins of the present invention may be combined with a pharmaceutically acceptable buffer, and the pH adjusted to provide acceptable stability, and a pH acceptable for administration such as parenteral administration.

Optionally, one or more pharmaceutically-acceptable antimicrobial agents may be added. Meta-cresol and phenol are preferred pharmaceutically-acceptable microbial agents. One or more pharmaceutically-acceptable salts may be added to adjust the ionic strength or tonicity. One or more excipients may be added to further adjust the isotonicity of the

- 62 -

formulation. Glycerin is an example of an isotonicity-adjusting excipient. Pharmaceutically acceptable means suitable for administration to a human or other animal and thus, does not contain toxic elements or undesirable 5 contaminants and does not interfere with the activity of the active compounds therein.

A pharmaceutically-acceptable salt form of the heterologous fusion proteins of the present invention may be used in the present invention. Acids commonly employed to 10 form acid addition salts are inorganic acids such as hydrochloric acid, hydrobromic acid, hydriodic acid, sulfuric acid, phosphoric acid, and the like, and organic acids such as *p*-toluenesulfonic acid, methanesulfonic acid, oxalic acid, *p*-bromophenyl-sulfonic acid, carbonic acid, succinic acid, 15 citric acid, benzoic acid, acetic acid, and the like.

Preferred acid addition salts are those formed with mineral acids such as hydrochloric acid and hydrobromic acid.

Base addition salts include those derived from inorganic bases, such as ammonium or alkali or alkaline earth metal 20 hydroxides, carbonates, bicarbonates, and the like. Such bases useful in preparing the salts of this invention thus include sodium hydroxide, potassium hydroxide, ammonium hydroxide, potassium carbonate, and the like.

25 Administration of Compositions:

Administration may be via any route known to be effective by the physician of ordinary skill. Peripheral, parenteral is one such method. Parenteral administration is commonly understood in the medical literature as the injection of a 30 dosage form into the body by a sterile syringe or some other mechanical device such as an infusion pump. Peripheral parenteral routes can include intravenous, intramuscular, subcutaneous, and intraperitoneal routes of administration.

The heterologous fusion proteins of the present invention 35 may also be amenable to administration by oral, rectal, nasal,

- 63 -

or lower respiratory routes, which are non-parenteral routes. Of these non-parenteral routes, the lower respiratory route and the oral route are preferred.

The fusion proteins of the present invention can be used 5 to treat a wide variety of diseases and conditions. The fusion proteins of the present invention primarily exert their biological effects by acting at a receptor referred to as the "GLP-1 receptor." Subjects with diseases and/or conditions that respond favorably to GLP-1 receptor stimulation or to the 10 administration of GLP-1 compounds can therefore be treated with the GLP-1 fusion proteins of the present invention. These subjects are said to "be in need of treatment with GLP-1 compounds" or "in need of GLP-1 receptor stimulation". Included are subjects with non-insulin dependent diabetes, 15 insulin dependent diabetes, stroke (see WO 00/16797), myocardial infarction (see WO 98/08531), obesity (see WO 98/19698), catabolic changes after surgery (see U.S. Patent No. 6,006,753), functional dyspepsia and irritable bowel syndrome (see WO 99/64060). Also included are subjects 20 requiring prophylactic treatment with a GLP-1 compound, e.g., subjects at risk for developing non-insulin dependent diabetes (see WO 00/07617). Subjects with impaired glucose tolerance or impaired fasting glucose, subjects whose body weight is about 25% above normal body weight for the subject's height 25 and body build, subjects with a partial pancreatectomy, subjects having one or more parents with non-insulin dependent diabetes, subjects who have had gestational diabetes and subjects who have had acute or chronic pancreatitis are at risk for developing non-insulin dependent diabetes.

30 An "effective amount" of a GLP-1 compound is the quantity which results in a desired therapeutic and/or prophylactic effect without causing unacceptable side-effects when administered to a subject in need of GLP-1 receptor stimulation. A "desired therapeutic effect" 35 includes one or more of the following: 1) an amelioration of

- 64 -

the symptom(s) associated with the disease or condition; 2) a delay in the onset of symptoms associated with the disease or condition; 3) increased longevity compared with the absence of the treatment; and 4) greater quality of life 5 compared with the absence of the treatment. For example, an "effective amount" of a GLP-1 compound for the treatment of diabetes is the quantity that would result in greater control of blood glucose concentration than in the absence of treatment, thereby resulting in a delay in the onset of 10 diabetic complications such as retinopathy, neuropathy or kidney disease. An "effective amount" of a GLP-1 compound for the prevention of diabetes is the quantity that would delay, compared with the absence of treatment, the onset of elevated blood glucose levels that require treatment with 15 anti-hypoglycaemic drugs such as sulfonyl ureas, thiazolidinediones, insulin and/or bisguanidines.

The dose of fusion protein effective to normalize a patient's blood glucose will depend on a number of factors, among which are included, without limitation, the subject's 20 sex, weight and age, the severity of inability to regulate blood glucose, the route of administration and bioavailability, the pharmacokinetic profile of the fusion protein, the potency, and the formulation.

The present invention comprises GLP-1 compounds that 25 have improved biochemical and biophysical properties by virtue of being fused to an albumin protein, an albumin fragment, an albumin analog, a Fc protein, a Fc fragment, or a Fc analog. These heterologous proteins can be successfully expressed in host cells, retain signaling 30 activities associated with activation of the GLP-1 receptor, and have prolonged half-lives.

The following examples are presented to further 35 describe the present invention. The scope of the present invention is not to be construed as merely consisting of the following examples. Those skilled in the art will recognize

- 65 -

that the particular reagents, equipment, and procedures described are merely illustrative and are not intended to limit the present invention in any manner.

5

Example 1: Construction of DNA

encoding heterologous fusion proteins

Example 1a Construction of DNA encoding Val⁸-GLP-1(7-37) -

10 Fc: A Fc portion of human IgG1 was isolated from a cDNA library and contains the full hinge region and the CH2 and CH3 domains. A fragment containing 696 base pairs of this Fc portion of human IgG1 was subcloned into the *NheI* and *Eco47III* sites of mammalian expression vector pJBO2 to create pJBO2/Fc (see Figure 5). DNA encoding the Igκ secretion signal sequence fused to Val⁸-GLP-1(7-37) was 15 generated by *in vitro* hybridization of four overlapping and complementary oligonucleotides:

5' - CTAGCCACC**ATGG**GAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTT
CCAGGTTCCACTGGTGACCAGTG - 3' [SEQ ID NO:12]

20

5' - GAGGGCACCTTCACCTCCGACGTGTCCTCCTATCTGGAGGGCCAGGCCGCCAAGGA
GTTCATCGCCTGGCTGGTGAAGGGAAGAGGC - 3' [SEQ ID NO:13]

25

5' - TGAAGGTGCCCTCCACGTGGTCACCACTGGAACCTGGAACCCAGAGCAGCAGTA
CCCATAGCAGGAGTGTGTCTGTCCATGGTGG - 3' [SEQ ID
NO:14]

30

The hybridization reaction was carried out using equivalent amounts of each oligonucleotide (1 pm/μl final concentration for each oligo). The mixture of oligonucleotides was heated for 5 min at 100°C in ligation buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 10 mM DTT, 1mM

- 66 -

ATP, 25 µg/ml bovine serum albumin) and then cooled over at least 2 hours to 30°C.

The resulting hybridization product was ligated for 2 hours at room temperature or overnight at 16°C to the

5 pJB02/Fc vector backbone which had been digested with *NheI* and *Eco47III*. The ligation products were used to transform competent XL-1 Blue cells (Stratagene). Recombinant plasmids were screened for the presence of peptide coding inserts by digesting clones with *NcoI* (encoding the Kozak 10 sequence and first Met of the signal peptide) and sequenced. The resulting expression plasmid used for transfection assays was denoted pJB02-V8-GLP-1-Fc (Figure 5).

Example 1b Construction of DNA encoding Val⁸-GLP-1(7-37)-

15 HSA: The plasmid HSA/pcDNA3.1GS was purchased from Invitrogen (Catalog # H-M12523M-pcDNA3.1/GS) and used as a template to isolate the cDNA encoding human serum albumin (HSA). The HSA cDNA was prepared using PCR wherein the DNA encoding the leader sequence as well as the six amino acid 20 pro-peptide was removed from the 5' end. In addition, stop codons were added directly at the 3' end of the HSA coding sequence. Finally, restriction enzyme sites were engineered at the 5' and 3' end to facilitate cloning. The HSA DNA sequence present in the original vector purchased from 25 Invitrogen contained a single base change in the 3' region of the gene (position 667) compared to the native human sequence. This change would result in a codon for Asn instead of Asp. Thus, using the strand overlapping PCR mutagenesis method discussed above, the codon was changed to 30 code for Asp at this position. The resulting HSA encoding DNA was cloned into the *NheI* and *HindIII* sites of pJB02 to create pJB02-HSA (Figure 6).

The Igκ leader sequence fused to the Val⁸-GLP-1(7-37) sequence was generated as discussed in Example 1a. This DNA

- 67 -

was ligated into the *NheI* and *FspI* sites of pJB02-HSA to create pJB02- Val⁸-GLP-1-HSA.

Example 1c Construction of DNA encoding Val⁸-GLP-1(7-37)-linker-HSA:

The vector pJB02-HSA was prepared as discussed in Example 1b. DNA encoding the linker sequence [GGGGS]₃ was ligated in frame to the 5' end of the HSA encoding DNA to create pJB02-linker-HSA (Figure 7). DNA encoding the IgK leader sequence and fused to the Val⁸-GLP-1(7-37) sequence and the 5' part of the linker sequence was generated as discussed in Example 1a. This DNA was ligated into the *NheI* and *BspEI* sites of pJB02 to create pJB02- Val⁸-GLP-1-linker-HSA.

Example 1d Construction of DNA encoding Exendin-4-Fc:

The plasmid pJB02/Fc was prepared as described in Example 1a. DNA encoding the IgK signal sequence fused to Exendin-4 was generated by *in vitro* hybridization of the following overlapping and complementary oligonucleotides:

5' - CTAGCCACCATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTG
GGTTCCAGGTTCCACCGGTCAC - 3' [SEQ ID NO:16]

5' - GGAGAGGGAACCTTCACCAGCGACCTGAGCAAGCAGATGGAGGAGGAGGCCGT
GAGACTG - 3' [SEQ ID NO:17]

5' - TTCATCGAGTGGCTGAAGAACGGAGGACCAAGCAGCGGAGCCCCTCCTCCT
AGC - 3' [SEQ ID NO:18]

5' - GAACCTGGAACCCAGAGCAGCAGTACCCATAGCAGGAGTGTGTCTGTCTCCA
TGGTGG - 3' [SEQ ID NO:19]

5' - CTCCTCCTCCATCTGCTTGCTCAGGTCGCTGGTGAAGGTTCCCTCTCCGTGA
CCGGTG - 3' [SEQ ID NO:20]

- 68 -

5' - 'GCTAGGAGGAGGGGCTCCGCTGCTGGTCCCTCCGTTCTTCAGCCACTCGAT
GAACAGTCTCACGGC - 3' [SEQ ID NO:21]

The hybridization reaction was carried out as described
5 in Example 1a. The hybridized product was ligated to the
pJB02 vector which had been digested with *NheI* and *Eco47III*
as described in Example 1a to create pJB02-Exendin-4-Fc.

Example 1e Construction of DNA encoding Exendin-4-HSA:

10 The plasmid pJB02-HSA was prepared as described in
Example 1b. DNA encoding the Ig κ signal sequence fused to
Exendin-4 was generated by *in vitro* hybridization of the
same overlapping and complementary oligonucleotides
described in Example 1d. Hybridization reactions were also
15 carried out as described above. DNA was cloned into unique
NheI and *FspI* sites in pJB02-HSA to create pJB02-Exendin-4-
HSA.

Example 1f Construction of DNA encoding Exendin-4-linker-
20 HSA:

The plasmid pJB02-linker-HSA was constructed as
described in Example 1c. DNA encoding the Ig κ signal
sequence fused to Exendin-4 and the 5' part of the linker
sequence was generated as in Example 1d. This DNA was
25 cloned into unique *NheI* and *BspEI* sites in pJB02-linker-HSA
to create pJB02-Exendin-4-linker-HSA.

Example 1g Construction of DNA encoding Val⁸-GLP-1/C-Ex-Fc:

The plasmid pJB02-Exendin-4-Fc was prepared as
30 described in Example 1d. The Exendin-4 encoding DNA was
excised from the vector with *AgeI* and *Eco47III*. The Val⁸-
GLP-1/C-Ex encoding DNA was generated by *in vitro*
hybridization of the following overlapping and complementary
oligonucleotides:

- 69 -

5' -CCGGTCACGTGGAGGGCACCTCACCTCCGACGTGTCCTCCTATCTGGA
GGGCCAGGCCGCCA - 3' [SEQ ID NO:22]

5' - AGGAATTCATCGCCTGGCTGGTGAAGGGCCGGGCAGCAGCGG
5 AGCCCCTCCTCCTAGC - 3' [SEQ ID NO:23]

5' - CTCCAGATAGGAGGACACGTCGGAGGTGAAGGTGCCCTCCAC
GTGA - 3' [SEQ ID NO:24]

10 5' - GCTAGGGAGGAGGGCTCCGCTGCTGCCCGGCCCTCACCA
GAGGCCA - 3' [SEQ ID NO:25]

The hybridization reaction was carried out as described
in Example 1a. The hybridized product was ligated in place
15 of Exendin-4 in the pJB02-Exendin-4-Fc expression vector to
create pJB02-Val⁸-GLP-1/C-Ex-Fc.

Example 1h Construction of DNA encoding Val⁸-Glu²²-GLP-1-Fc:

The plasmid pJB02-Exendin-4-Fc was prepared as
20 described in Example 1d. The Exendin-4 encoding DNA was
excised from the vector with *AgeI* and *Eco47III*. The Val⁸-
Glu²²-GLP-1 encoding DNA was generated by *in vitro*
hybridization of the following overlapping and complementary
oligonucleotides:

25 5' -CCGGTCACGTGGAGGGCACCTCACCTCCGACGTGTCCTCCTATCTGGA
GGAGCAGGCCGCCA - 3' [SEQ ID NO:26]

30 5' - AGGAGTTCATCGCCTGGCTGGTGAAGGGCCGGGC - 3' [SEQ ID NO:27]

5' - GCCCCGGCCCTTCACCAGCCAGGCGATGAACTCCTTGGCGGCC
TGCTC - 3' [SEQ ID NO:28]

35 5' - CTCGAGATAGGAGGACACGTCGGAGGTGAAGGTGCCCT
CCACGTGA - 3' [SEQ ID NO:29]

- 70 -

The hybridization reaction was carried out as described in Example 1a. The hybridized product was ligated in place of Exendin-4 in the pJB02-Exendin-4-Fc expression vector to 5 create pJB02-Val⁸-Glu²²-GLP-1-Fc.

Example 1i Construction of DNA encoding Val⁸-Glu²²GLP-1/C-Ex-Fc:

The plasmid pJB02-Exendin-4-Fc was prepared as 10 described in Example 1d. The Exendin-4 encoding DNA was excised from the vector with *AgeI* and *Eco47III*. The Val⁸-Glu²²GLP-1/C-Ex encoding DNA was generated by *in vitro* hybridization of the following overlapping and complementary oligonucleotides:

15

5' - CCGGTCACGTGGAGGGCACCTCACCTCCGACGTGTCCTCCTATCTCGA
GGAGCAGGCCGCCA - 3' [SEQ ID NO:30]

20

5' - AGGAATTCATCGCCTGGCTGGTGAAGGGCCGGGCAGCAGCGGA
GCCCTCCTCCTAGC - 3' [SEQ ID NO:31]

5' - CTCGAGATAGGAGGACACGTCGGAGGTGAAGGTGCC
TCCACGTGA - 3' [SEQ ID NO:32]

25

5' - GCTAGGAGGAGGGGCTCCGCTGCTGCCCGGCCCTCACCA
GAGCGA
TGAATTCCCTTGGCGGCCCTGCTC - 3' [SEQ ID NO:33]

30

The hybridization reaction was carried out as described in Example 1a. The hybridized product was ligated in place of Exendin-4 in the pJB02-Exendin-4-Fc expression vector to create pJB02-Val⁸-Glu²²-GLP-1/C-Ex-Fc.

Example 1j Construction of DNA encoding Gly⁸-GLP-1-Fc:

The plasmid pJB02-Exendin-4-Fc was prepared as 35 described in Example 1d. The Exendin-4 encoding DNA was

- 71 -

excised from the vector with *AgeI* and *Eco47III*. The Gly⁸-GLP-1 encoding DNA was generated by *in vitro* hybridization of the following overlapping and complementary oligonucleotides:

5

5' - CCGGTCACGGCGAGGGCACCTCACTAGTGACGTGTCCTCCTATCTGGA
GGGCCAGGCCGCCA - 3' [SEQ ID NO:34]

5' - AGGAGTTCATCGCCTGGCTGGTGAAGGGCCGGGC - 3' [SEQ ID NO:35]

10

5' - CTCCAGATAGGAGGACACGTCACTAGTGAAAGGTGCCCTC
GCCGTGA - 3' [SEQ ID NO:36]

5' - GCCCCGGCCCTTCACCAGCCAGGCGATGAACTCCTGGCGGC
15 CTGGCC - 3' [SEQ ID NO:37]

20

The hybridization reaction was carried out as described in Example 1a. The hybridized product was ligated in place of Exendin-4 in the pJB02-Exendin-4-Fc expression vector to create pJB02-Gly⁸-GLP-1-Fc.

Example 2: Expression of heterologous fusion proteins

Expression of the fusion proteins encoded by the DNA constructs of Example 1 was carried out by transiently transfecting HEK 293EBNA cells (both adherent and suspension). Cells were counted and seeded 24 hours prior to transfection. The transfection cocktail was prepared by mixing FuGeneTM6 transfection reagent (Roche Molecular Biochemicals, catalog # 1814443) with OptiMEM (Gibco/BRL) and incubating at room temperature for 5 min at which point DNA was added and the cocktail was incubated for an additional 15 min. Immediately before transfection, fresh growth media was added to the plate. Tables 1 and 2 provide further transfection details.

35

- 72 -

Table 1: Reagents used in transient transfections of 293EBNA cells.

Tissue culture vessel	Number of cells seeded	DNA (μg)	FuGene (μl)	OptiMEM media (ml)	Vol. of growth medium (ml)
35 mm	5 X 10 ⁵	1.5	9	0.102	2
100 mm	2 X 10 ⁶	12	73	0.820	10
700 cm ² (RB)	2 X 10 ⁷	65	400	4.0	100

5 Table 2: Media composition

Growth and transfection medium	Harvesting medium
DMEM F12 3:1	Hybritech base
5 % FBS	1 mM Ca ²⁺
20 mM HEPES	20 mM HEPES
2 mM L-glutamine	1 μg/ml Nuselin (human insulin)
50 μg/ml geneticin (G418 NEO)	1 μg/ml human transferrin
50 μg/ml tobromycin	50 μg/ml tobromycin

For small-scale transfections (35mm - 10mm vessels), cells were rinsed with PBS and switched to harvesting media 10 24 hours post- transfection and media was collected and replaced every 24 hours for several days. In the case of large-scale transfections (700 cm² roller bottles), the roller bottles were rinsed with PBS 48 hours post-transfection and changed to harvesting media. Media was 15 collected and changed every 24 hours for at least 10 consecutive days. Routinely, only 10 harvests were used for subsequent protein purification.

- 73 -

Example 3: Purification of heterologous fusion proteins

Example 3a purification of Val⁸-GLP-1-Fc

Approximately 4.5 liters of conditioned medium (fusion protein expression level approximately 20 μ g/ml) from large-
5 scale transfections was filtered using a CUNO filter system and concentrated to 250 ml using a ProFlux tangential flow filtration system with a 10 K filter membrane. Val⁸-GLP-1-Fc was captured with a 5 ml HiTrap protein A column in 1x PBS, pH 7.4 at a flow rate of 2 ml/min and eluted with 50 mM
10 citric acid pH 3.3. Fractions (1 ml) were collected in tubes containing 4 ml 1x PBS and 100 μ l 1 M Tris pH 8.

Fractions containing the fusion protein, as determined by SDS-PAGE and reverse phase-HPLC on Zorbax C8, were pooled and applied to a Superdex 75 60/60 column in 1x PBS pH 7.4
15 at a flow rate of 10 ml/min. Positive fractions (20 mls/tube) were collected and pooled. Pooled fractions were then subjected to C4 Reverse Phase Chromatography in 0.1 %TFA water at a flow rate of 3 ml/min. Val⁸-GLP-1-Fc was eluted using a gradient from 5% B (0.1% TFA in acetonitrile)
20 to 100% B in 70 min. Eluant fractions (3 mls/tube) were collected. Acetonitrile was removed by vacuum drying and 1 ml of H₂O was added. The purified sample (approximately 32 mls) was dialyzed twice against 4 liters of 1x PBS pH7.4.

The dialyzed sample was then filtered using a MILLEX-GV 25 0.22 um Filter Unit and concentration was determined using absorption at 280 nm.

Example 3b purification of Val⁸-GLP-1-HSA or Val⁸-GLP-1-Linker-HSA

30 Approximately 6.5 liters of conditioned medium (fusion protein expression level approximately 10 μ g/ml) was filtered using a CUNO filter system and concentrated to 380 mls using a ProFlux tangential flow filtration system with a 10 K filter membrane.

- 74 -

The fusion protein was captured using a 50 ml Fast Flow Q column (Pharmacia) in 20 mM Tris pH 7.4 at a flow rate of 5ml/min. Protein was eluted using a gradient: from 0% to 50% 20mM Tris pH 7.4, 1M NaCl in 10 CV, then to 100%B in 2 CV.

5 Fractions containing the fusion protein were pooled and subjected to C4 Reverse Phase Chromatography in 0.1% TFA water at a flow rate of 5 ml/min. The fusion protein was eluted using a gradient from 20% B (0.1% TFA in acetonitrile) to 90% B in 120min. Fractions (3.5 ml/tube) 10 were collected. Acetonitrile was removed by vacuum drying.

Approximately 9 mls of pooled sample was diluted with 1x PBS pH 7.4 to 40ml and dialyzed against 4 liters of 1x PBS pH 7.4 overnight. The sample was filtered and concentration was determined by absorption at 280nm.

15

Example 3c purification of Exendin-4-Fc:

Approximately 4 liters of conditioned medium (fusion protein expression level approximately 8 μ g/ml) was filtered using a CUNO filter system and concentrated to 250 mls using 20 a ProFlux tangential flow filtration system with a 30K filter membrane.

Exendin-4-Fc was captured with a 5 ml HiTrap protein A column in 1x PBS, pH 7.4 at a flow rate of 2 ml/min and eluted with 50 mM citric acid pH 3.3. Fractions containing 25 the fusion protein were pooled, filtered, and dialyzed against 4 liters of 1 x PBS over night. The dialyzed sample was then applied to a Superdex 75 60/60 column in 1x PBS pH7.4, 0.5M NaCl at a flow rate of 10 ml/min. Fractions (20 ml/tube) containing the fusion protein were collected, 30 pooled, and concentrated to about 1 mg/ml. Concentrated samples were then filtered using a MILLEX-GV 0.22 um Filter Unit.

- 75 -

Example 3d purification of Exendin-4-HSA and Exendin-4-linker-HSA:

5 Approximately 1.1 liters of conditioned medium (fusion protein expression level approximately 6 μ g/ml) was filtered using a CUNO filter system and concentrated to 175 mls using a ProFlux tangential flow filtration system with a 30K filter membrane.

10 The fusion protein was captured using a 5 ml HiTrap Q-sepharose column (Pharmacia) in 20 mM Tris pH 7.4 at a flow rate of 2 ml/min. Protein was eluted using a gradient from 0% to 50% 20mM Tris pH 7.4, 1M NaCl in 12 CV and then to 100%B in 4 CV.

15 Fractions containing the fusion protein were pooled and subjected to C4 Reverse Phase Chromatography in 0.1% TFA water at a flow rate of 5 ml/min. The fusion protein was eluted using a gradient from 10% B (0.1% TFA in acetonitrile) to 100% B in 70 min. Fractions (10 ml/tube) containing the fusion protein were collected. Acetonitrile 20 was removed using a vacuum dryer.

25 Approximately 8 mls of pooled sample was dialyzed against 4 liters of 1x PBS pH 7.4 overnight. The sample was filtered and concentration was determined by absorption at 280nm. The dialyzed sample was then applied to a Superdex 200 26/60 column in 1x PBS pH 7.4, 0.5 M NaCl at a flow rate of 2 ml/min. Fractions (3 ml/tube) containing the fusion protein were collected, pooled, concentrated, and filtered.

Example 4: Characterization of fusion proteins by SDS PAGE:

30 SDS-PAGE followed by immunoblotting was used to analyze both purified fusion protein as well as conditioned medium from cells transfected with various fusion protein expression vectors. SDS-PAGE was performed on a Novex Powerease 500 system using Novex 16% Tris-Glycine Precast 35 gels (EC6498), running buffer (10x, LC2675) and sample

- 76 -

buffer (L2676). Samples were reduced with 50 mM DTT and heated 3-5 min at 95°C prior to loading.

After running the SDS-PAGE gel, water and transfer buffer (1X Tris-Glycine Seprabuff (Owl Scientific Cat. No. 5 ER26-S) with 20% methanol) were used to rinse SDS from the gels. A Novex transfer apparatus was used with PVDF (BioRad, Cat. No. 162-0174) and nitrocellulose membranes (BioRad, Cat. No. 1703965 or 1703932). Transfer was carried out at room temperature for 90 min at 30-35 V. Membranes were 10 blocked in 1X PBS with 0.1% Tween-20 (Sigma, Cat. No. P-7949) and 5% Milk (BioRad, Cat. No. 170-6404) for 1-12 hours at 4°C. Antibodies are diluted into 1X PBS +5% Milk and the blots are incubated in these solutions for 1-2 h at 4°C. Between incubations, the blots are washed four times for 15 min each with 1X PBS and 0.2% Tween-20 at room temperature. PBS was made from either GIBCO 10X PBS (Cat No. 70011), to give a final composition of 1 mM monobasic potassium phosphate, 3 mM dibasic sodium phosphate, 153 mM sodium chloride, pH 7.4, or PBS pouches from Sigma (Cat. No. 1000-20 3), to give 120 mM NaCl, 2.7 mM KCl and 10 mM phosphate, pH 7.4 at 25°C.

The primary antibodies were either a polyclonal goat anti-IgG1 or rabbit anti-HSA. The secondary antibody was either an anti-goat IgG HRP or an anti-rabbit IgG HRP. The 25 secondary antibody was diluted 1:5000. An ECL system (Amersham Pharmacia Biotech, Cat. No. RN2108 and Cat. No. RPN1674H) was used for developing blots.

Figure 3A compares purified Fc protein to conditioned media from pJB02-⁸Val-GLP-1-Fc and pJB02-Exendin-4-Fc 30 transfected cells. The decrease in mobility is consistent with the increased size due to the GLP-1 portion of the fusion protein. Figure 3B similarly compares purified HSA with conditioned media from cells transfected with pJB02-⁸Val-⁸GLP-2-HSA, pJB02-⁸Val-GLP-1-Linker-HSA, pJB02-Exendin-

- 77 -

4-HSA, or pJB02-Exendin-4-Linker-HSA. Figure 4 identifies purified fusion protein preparations.

Example 5: Characterization of fusion
5 proteins using mass spectrometry:

All experiments were performed on a Micromass TofSpec-2E mass spectrometer equipped with Time Lag Focusing electronics, a Reflectron (used in analysis of the 0-8000 Da peptide range), a Linear detector (used during high mass / 10 good signal analysis), and Post Acceleration Detector (or P.A.D., used for high mass / extremely low signal analysis) The effective path length of the instrument in Linear mode is 1.2 meters, in Reflectron mode it is 2.3 meters. Two 15 dual micro-channel plate detectors are fitted for linear and reflectron mode detection. The laser used is a Laser Science Inc. VSL-337i nitrogen laser operating at 337 nm at 5 laser shots per second. All data were acquired using a 2 GHz, 8 bit internal digitizer and up to 50 laser shots were averaged per spectrum.

20 The instrument was operated in linear mode for the analysis of the GLP-1 fusion proteins in question. The linear detector is a device that detects ions that travel down the flight tube of the MALDI-ToF-MS instrument. It measures the ion abundance over time and sends a signal to 25 the digitizer for conversion. The digitizer is an analog-to-digital converter that allows the signal from the mass spectrometer to be transferred to the computer, where it is reconstructed into a usable m/z spectrum.

30 A recrystallized saturated sinapinic acid solution (diluted in 50/50 Acn / H₂O and 0.1% TFA) was utilized as the ionization matrix. Sinapinic acid is a proper matrix for proteins above 10 kDa. Mass appropriate reference proteins were used for internal and external calibration files in order to obtain accurate mass determinations for the samples 35 analyzed. Samples were all analyzed using a 1:2 sample to

- 78 -

matrix dilution. The instrument was initially set up under the following linear detector conditions:

Source Voltage: 20.0 keV **Pulse Voltage:** 3.0 keV
Extraction Voltage: 20.0 keV **Laser Coarse:** 50
Focus Voltage: 16.0 keV **Laser Fine:** 50
Linear detector: 3.7 keV
P.A.D.: (off line)

10 These settings were modified (as needed) to give the best signal/noise ratio and highest resolution. Table 3 provides a characterization of different GLP-1 fusion proteins.

15 **Table 3**

Fusion Protein	Expected Mass (kDa)	Mass determined by Mass Spec (kDa)
Val ⁸ -GLP-1-IgG1	59.08	61.94
Val ⁸ -Glu ²² -GLP-1-IgG1	59.23	63.61
Gly ⁸ -GLP-1-IgG1	59.00	62.93
Val ⁸ -GLP-1-CEx-IgG1	60.45	65.1-65.6
Val ⁸ -Glu ²² -GLP-1-CEx-IgG1	60.69	65.86
Exendin-4-IgG1	60.69	65.86
Val ⁸ -GLP-1-Linker-HSA	70.70	69.89, 70.74
Exendin-4-HSA	70.56	70.62
Exendin-4-Linker-HSA	71.56	71.62

CEx refers to a C-terminal extension and comprises the sequence of Ser-Ser-Gly-Ala-Pro-Pro-Ser.
 Linker is Gly-Gly-Gly-Gly-Ser-Gly-Gly-Gly-Gly-Ser-Gly-Gly-Gly-Ser.

- 79 -

Example 6: Activity of heterologous fusion proteins:

The ability of the fusion proteins of the present invention to activate the GLP-1 receptor was assessed using *in vitro* assays such as those described in EP 619,322 to 5 Gelfand, *et al.*, and U.S. Patent No. 5,120,712, respectively. The activity of these compounds relative to the activity of Val⁸-GLP-1(7-37)OH is reported in Table 4. Figure 8 represents *in vitro* dose response curves for Val⁸-GLP-1 and Exendin-4 fusion proteins. In addition, Table 5a 10 and 5b provide the *in vitro* activity of a large group of GLP-1 analogs that can be fused to an Fc or an albumin protein to make biologically active fusion proteins. These activities are compared to GLP-1(7-37)OH.

15 Table 4: *In vitro* activity of GLP-1 fusion proteins

Fusion Protein	In Vitro Activity (% of Val ⁸ -GLP-1)
Val ⁸ -GLP-1-IgG1	1
Exendin-4-IgG1	240
Val ⁸ -GLP-1-Linker-HSA	0.2
Exendin-4-HSA	20
Exendin-4-Linker-HSA	90
Exendin-4	500
Val ⁸ -Glu ²² -GLP-1-IgG1	3.7
Gly ⁸ -GLP-1-IgG1	3.3
Val ⁸ -GLP-1-CEx-IgG1	3.3
Val ⁸ -Glu ²² -GLP-1-CEx-IgG1	29
Gly ⁸ -Glu ²² -GLP-1-C2-IgG1	75
Gly ⁸ -Glu ²² -GLP-1-CEx-Linker-IgG1	150
Exendin-4-C2-IgG1	250
Exendin-4-Linker-IgG1	330
Gly ⁸ -Glu ²² -GLP-1-CEx-Linker-HSA	4
Gly ⁸ -Glu ²² -GLP-1-CEx-Linker-IgG4	80

- 80 -

CEx refers to a C-terminal extension and comprises the sequence of Ser-Ser-Gly-Ala-Pro-Pro-Pro-Ser.
Linker is Gly-Gly-Gly-Gly-Ser-Gly-Gly-Gly-Ser-Gly-Gly-Gly-Ser

5 C2 is Ser-Ser-Gly-Ala-Ser-Ser-Gly-Ala.

The amino acid sequences of the fusion proteins described in Tables 3 and 4 are represented in SED ID NO: 13 to SEQ ID NO: 31.

Val⁸-GLP-1-Human serum albumin amino acid sequence is

10 represented by SEQ ID NO: 13.

```

1 HVEGTFTSDV SSYLEGQAAK EFIGLVLKGR GDAHKSEVAH RFKDLGEENF KALVLIAFAQ
61 YLQQCPFEDH VKLVNEVTEF AKTCVADESA ENCDKSLHTL FGDKLCTVAT LRETYGEMAD
121 CCAKQEPEERN ECFLQHKDDN PNLPRLVRPE VDVMCTAFHD NEETFLKKYL YEIARRHYPF
181 YAPELLFFAK RYKAATTECC QAADKAACLL PKLDELRDEG KASSAKQRLK CASLQKFGER
15 241 AFKAWAVARL SQRFPKAEEFA EVSKLVTDLT KVHTECCHGD LLECADDRAD LAKYICENQD
301 SISSKLKECC EKPLLEKSHC IAEVENDEMP ADLPSLAADF VESKDVKNY AEAKDVFLGM
361 FLYEYARRHP DYSVVVLLRL AKTYETTLEK CCAAADPHEC YAKVDEFKP LVEEPQNLIK
421 QNCELFEQLG EYKFQNALLV RYTKKVPQVS TPTLVEVSRN LGKVGSKCCK HPEAKRMPCA
481 EDYLSVVLNQ LCVLHEKTPV SDRVTKCCTE SLVNRRPCFS ALEVDETYVP KEFNAETFTF
20 541 HADICTLSEK ERQIKKQTAL VELVKHKPKA TKEQLKAVMD DFAAFVEKCC KADDKETCFA
601 EEGKKLVAAS QAALGL [SEQ ID NO: 13]
```

Val⁸-GLP-1-Linker-Human serum albumin amino acid sequence is represented by SEQ ID NO: 14.

```

25 1 HVEGTFTSDV SSYLEGQAAK EFIGLVLKGR GGGGGSGGGG SGGGGSDAHK SEVAHRFKDL
61 GEENFKALVL IAFAYLQQC PFEDHVVLN EVTEFAKTCV ADESAENCDK SLHTLFGDKL
121 CTVATLRETY GEMADCCAQ EPERNECFLQ HKDDNPNLPR LVRPEVDVMC TAFHDNEETF
181 LKKYLYEIAR RHPYFYAPEL LFFAKRYKAA FTECCQAADK AACLLPKLDE LRDEGKASSA
241 KQRLKCASLQ KFGERAFKAW AVARLSQRFP KAEFAEVSKL VTDLTKVHTE CCHGDLLECA
30 301 DDRADLAKYI CENQDSISSLK LKECCEKPLL EKSHCIAEVE NDEMPADLPS LAADFVESKD
361 VCKNYAEAKD VFLGMFLYEEY ARRHPDYSVV LLLRLAKTYE TTLEKCCAAA DPHECYAKVF
421 DEFKPLVEEP QNLIKQNCEL FEQLGEYKFQ NALLVRYTKK VPQVSTPTLV EVSRNLGKVG
481 SKCCKHPEAK RMPCAEDYLS VVLNQLCVLH EKTPVSDRVT KCCTESLVNR RPCFSALEVD
541 ETYVPKEFNA ETFTFHADIC TLSEKERQIK KQTALVELVK HKPKATKEQL KAVMDDFAAF
35 601 VEKCKADDK ETCFAEKGK LVAASQAALG L [SEQ ID NO: 14]
```

- 81 -

Gly⁸-Glu²²-GLP-1-CEx-Linker-Human serum albumin amino acid sequence is represented by SEQ ID NO: 15.

1 HGEGTFTSDV SSYLEEQAAK EFIAWLVKGR GSSGAPPPSG GGGGSGGGGS GGGGSDAHKS
 61 EVAHRFKDLG EENFKALVLI AFAQYLQQCP FEDHVKLVNE VTEFAKTCVA DESAENCDKS
 5 121 LHTLFGDKLC TVATLRETYG EMADCCAKQE PERNECFLQH KDDNPNLPRl VRPEVDVMCT
 181 AFHDNEETFL KKYLYEIARR HPYFYAPELL FFAKRYKAaf TECCQAADKA ACLLPKLDEL
 241 RDEGKASSAK QRLKCASLQK FGERAFKAWA VARLSQRFPK AEFAEVSKLV TDLTKVHTEC
 301 CHGDLLECAD DRADLAKYIC ENQDSISSKL KECCEKPLLE KSHCIAEVEN DEMPADLPSL
 361 AADFVESKDV CKNYAEAKDV FLGMFLYELYA RRHPDYSVVL LLRLAKTYET TLEKCAAAD
 10 421 PHECYAKVFD EFKPLVEEPQ NLIKQNCELF EQLGEYKFQN ALLVRYTKKV PQVSTPTLVE
 481 VSRNLGKVGS KCCKHPEAKR MPCAEODYLSV VLNQLCVLHE KTPVSDRVTK CCTESLVNRR
 541 PCFSALEVDE TYVPKEFNAE TFTFHADICT LSEKERQIKK QTALVELVKH KPKATKEQLK
 601 AVMDDFAAFV EKCKADDKE TCFAEEGKKL VAASQAALGL [SEQ ID NO: 15]

15 Exendin-4-Human serum albumin amino acid sequence is represented by SEQ ID NO: 16.

1 HGEGTFTSDL SKQMEEEAVR LFIEWLKNGG PSSGAPPPSD AHKSEVAHRF KDLGEENFKA
 61 LVLIQAQYL QQCPFEDHVK LVNEVTEFAK TCVADESAEN CDKSLHTLFG DKLCTVATLR
 121 ETYGEMADCC AKQEPEPERNEC FLQHKDDNPN LPRLVPEVD VMCTAFHDNE ETFLKKYLYE
 20 181 IARRHPYFYA PELLFFAKRY KAAFTECCQA ADKAACLLPK LDELRDEGKA SSAKQRLKCA
 241 SLQKFGERAf KAWAVARLSQ RFPKAЕFAEV SKLVTDLTKV HTECCHGDLL ECADDRADLA
 301 KYICENQDSI SSKLKECCEK PLLEKSHCIA EVENDEMPAD LPSLAADFVE SKDVCKNYAE
 361 AKDVFGLMFL YEYARRHPDY SVVLLRLAK TYETTLEKCC AAADPHECYA KVFDEFKPLV
 421 EEPQNLIKQN CELFEQLGEY KFQNALLVRY TKKVPQVSTP TLVEVSRNLG KVGSKCKHP
 25 481 EAKRMPCAED YLSVVLNQLC VLHEKTPVSD RVTKCCTESL VNRRPCFSAL EVDETYVPKE
 541 FNAETFTFHA DICTLSEKER QIKKQTALVE LVKKHPKATK EQLKAVMDDF AAFVEKCKA
 601 DDKETCFAEE GKKLVAASQA ALGL [SEQ ID NO: 16]

Exendin-4-Linker-Human serum albumin amino acid sequence is represented by SEQ ID NO: 17.

1 HGEGTFTSDL SKQMEEEAVR LFIEWLKNGG PSSGAPPPSG GGGGSGGGGS GGGGSDAHKS
 61 EVAHRFKDLG EENFKALVLI AFAQYLQQCP FEDHVKLVNE VTEFAKTCVA DESAENCDKS
 121 LHTLFGDKLC TVATLRETYG EMADCCAKQE PERNECFLQH KDDNPNLPRl VRPEVDVMCT
 181 AFHDNEETFL KKYLYEIARR HPYFYAPELL FFAKRYKAaf TECCQAADKA ACLLPKLDEL
 35 241 RDEGKASSAK QRLKCASLQK FGERAFKAWA VARLSQRFPK AEFAEVSKLV TDLTKVHTEC
 301 CHGDLLECAD DRADLAKYIC ENQDSISSKL KECCEKPLLE KSHCIAEVEN DEMPADLPSL
 361 AADFVESKDV CKNYAEAKDV FLGMFLYELYA RRHPDYSVVL LLRLAKTYET TLEKCAAAD
 421 PHECYAKVFD EFKPLVEEPQ NLIKQNCELF EQLGEYKFQN ALLVRYTKKV PQVSTPTLVE

- 82 -

481 VSRNLGKVGS KCCKHPEAKR MPCAEODYLSV VLNQLCVLHE KTPVSDRVTK CCTESLVNRR
541 PCFSALEVDE TYVPKEFNAE TFTFHADICT LSEKERQIKK QTALVELVKH KPKATKEQLK
601 AVMDDFAAFV EKCKKADDKE TCFAEEGKKL VAASQAALGL [SEQ ID NO: 17]

5 Val^8 -GLP-1-IgG1 amino acid sequences represented by SEQ ID NO: 18.

1 HVEGTFTSDV SSYLEGQAAK EFIAWLVKGR GAEPKSCDKT HTCPPCPAPE LLGGPSVFLF
61 PPKPKDTLMI SRTPEVTCVV VDVSHEDEPV KFNWYVDGVE VHNAKTKPRE EQYNSTYRVV
121 SVLTVLHQDW LNGKEYKCKV SNKALPAPIE KTISKAKGQP REPQVYTLPP SREEMTKNQV
10 181 SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDSDGS FFLYSKLTVD KSRWQQGNVF
241 SCSVMHEALH NHYTQKSLSL SPGK [SEQ ID NO: 18]

Val^8 -GLP-1-Cex-IgG1 amino acid sequence is represented by SEQ ID NO: 19.

15 1 HVEGTFTSDV SSYLEGQAAK EFIAWLVKGR GSSGAPPSA EPKSCDKTHT CPPCPAPEELL
61 GGPSVFLFPP KPKDTLMISR TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ
121 YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN KALPAPIEKT ISKAKGQP PRE PQVYTLPPSR
181 EEMTKNQVSL TCLVKGFYPS DIAVEWESNG QPENNYKTTP PVLDSDGSFF LYSKLTVDKS
241 RWQQGNVFSC SVMHEALHNH YTQKSLSLSP GK [SEQ ID NO: 19]

20 Val^8 -Glu²²GLP-1-IgG1 amino acid sequence is represented by SEQ ID NO: 20.

1 HVEGTFTSDV SSYLEEQAAK EFIAWLVKGR GAEPKSCDKT HTCPPCPAPE LLGGPSVFLF
61 PPKPKDTLMI SRTPEVTCVV VDVSHEDEPV KFNWYVDGVE VHNAKTKPRE EQYNSTYRVV
25 121 SVLTVLHQDW LNGKEYKCKV SNKALPAPIE KTISKAKGQP REPQVYTLPP SREEMTKNQV
181 SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDSDGS FFLYSKLTVD KSRWQQGNVF
241 SCSVMHEALH NHYTQKSLSL SPGK [SEQ ID NO: 20]

30 Val^8 -Glu²²GLP-1-CEx-IgG1 amino acid sequence is represented by SEQ ID NO: 21.

1 HVEGTFTSDV SSYLEEQAAK EFIAWLVKGR GSSGAPPSA EPKSCDKTHT CPPCPAPEELL
61 GGPSVFLFPP KPKDTLMISR TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ
121 YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN KALPAPIEKT ISKAKGQP PRE PQVYTLPPSR
181 EEMTKNQVSL TCLVKGFYPS DIAVEWESNG QPENNYKTTP PVLDSDGSFF LYSKLTVDKS
35 241 RWQQGNVFSC SVMHEALHNH YTQKSLSLSP GK [SEQ ID NO: 21]

- 83 -

Gly⁸-Glu²²GLP-1-C2-IgG1 amino acid sequence is represented by SEQ ID NO: 22.

1 HGEGTFTSDV SSYLEEQAAK EFIAWLVKGR GSSGASSGAA EPKSCDKTHT CPPCPAPELL
61 GGPSVFLFPP KPKDTLMISR TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ
5 121 YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN KALPAPIEKT ISKAKGQPRES PQVYTLPPSR
181 EEMTKNQVSL TCLVKGFYPS DIAVEWESNG QPENNYKTTP PVLDSDGSFF LYSKLTVDKS
241 RWQQGNVFSC SVMHEALHNH YTQKSLSLSP GK [SEQ ID NO: 22]

Gly⁸-Glu²²GLP-1-CEx-Linker-IgG1 amino acid sequence is represented by SEQ ID NO: 23.

1 HGEGTFTSDV SSYLEEQAAK EFIAWLVKGR GSSGAPPPSG GGGGGGGSG GGGSAEPKSC
61 DKTHTCPPCP APELLGGPSV FLFPPPKPKDT LMISRTPEVT CVVVDVSHED PEVKFNWYVD
121 GVEVHNNAKTK PREEQYNSTY RVVSVLTVLH QDWLNGKEYK CKVSNKALPA PIEKTISKAK
181 GQPREPQVYT LPPSREEMTK NQVSLTCLVK GFYPSDIAVE WESNGQPENN YKTPPVLDLS
15 241 DGSFFFLYSKL TVDKSRWQQG NVFSCSVMHE ALHNHYTQKS LSLSPGK [SEQ ID NO: 23]

Gly⁸-Glu²²GLP-1-CEx-Linker-IgG4 amino acid sequence is represented by SEQ ID NO: 24.

1 HGEGTFTSDV SSYLEEQAAK EFIAWLVKGR GSSGAPPPSG GGGGGGGSG GGGSAESKYD
20 61 PPCPSCPAPPE FLGGPSVFLF PPCKDTLMI SRTPEVTCVV DVDSQEDPEV QFNWYVDGVE
121 VHNAKTKPRES EQFNSTYRVV SVLTVLHQDWLN LNGKEYKCKV SNKGLPSSIE KTISKAKGQP
181 REPQVYTLPP SQEEMTKNQV SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDSDGS
241 FFLYSRLTVD KSRWQEGNVF SCSVMHEALH NHYTQKSLSL SLGK [SEQ ID NO: 24]

25 Gly⁸-Glu²²GLP-1-CEx-2Linker-IgG1 amino acid sequence is represented by SEQ ID NO: 25.

1 HGEGTFTSDV SSYLEEQAAK EFIAWLVKGR GSSGAPPPSG GGGGGGGSG GGGGGGGSG
61 GGGGGGGSA EPKSCDKTHT CPPCPAPELL GGPSVFLFPP KPKDTLMISR TPEVTCVVVD
121 VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN
30 181 KALPAPIEKT ISKAKGQPRES PQVYTLPPSR EEMTKNQVSL TCLVKGFYPS DIAVEWESNG
241 QPENNYKTTP PVLDSDGSFF LYSKLTVDKS RWQQGNVFSC SVMHEALHNH YTQKSLSLSP
301 GK [SEQ ID NO: 25]

35 Gly⁸-Glu²²GLP-1-2Linker-IgG1 amino acid sequence is represented by SEQ ID NO: 26.

- 84 -

1 HGEGTFTSDV SSYLEEQAAK EFIAWLVKGR GGGGGSGGGG SGGGGSGGGG SGGGGSGGGG
61 SAEPKSCDKT HTCPCPAPE LLGGPSVFLF PPKPKDTLMI SRTPEVTCVV DVSHEDPEV
121 KFNWYVDGVE VHNAKTKPRE EQYNSTYRVV SVLTVLHQDW LNGKEYKCKV SNKALPAPIE
181 KTISKAKGQP REPQVYTLPP SREEMTKNQV SLTCLVKGFY PSDIAVEWES NGQPENNYKT
5 241 TPPVLDSDGS FFLYSKLTVD KSRWQQGNVF SCSVMHEALH NHYTQKSLSL SPGK
[SEQ ID NO: 26]

Gly⁸-Glu²²GLP-1-2CEx-IgG1 amino acid sequence is represented
by SEQ ID NO: 27.

10 1 HGEGTFTSDV SSYLEEQAAK EFIAWLVKGR GSSGAPPPSS SGAPPSAEP KSCDKTHTCP
61 PCPAPELLGG PSVFLFPPKP KDTLMISRTP EVTCVVVDVS HEDPEVKFNW YVDGVEVHNA
121 KTKPREEQYN STYRVSVLTL VHLDWLNGK EYKCKVSNKA LPAPIEKTIS KAKGQPREGQ
181 VYTLPPSREE MTKNQVSLTC LVKGFPYPSDI AVEWESNGQP ENNYKTTPPV LDSDGSFFLY
241 SKLTVDKSRW QQGNVFSC CSV MHEALHNHYT QKSLSLSPGK [SEQ ID NO: 27]

15 Gly⁸-Glu²²-Val²⁵-Ile³³GLP-1-CEx-Linker-IgG1 amino acid
sequence is represented by SEQ ID NO: 28.

1 HGEGTFTSDV SSYLEEQAVK EFIAWLIKGR GSSGAPPPSG GGGSGGGSG GGGSAEPKSC
61 DKTHTCPPCP APELLGGPSV FLFPPKPDT LMISRTPEVT CVVVDVSHED PEVKFNWYVD
20 121 GVEVHNAKTK PREEQYNSTY RVVSVLTVLH QDWLNGKEYK CKVSNKALPA PIEKTISKAK
181 GQPREGQVYT LPPSREEMTK NQVSLTCLVK GFYPSDIAVE WESNGQPENN YKTTPPVLD
241 DGSFFFLYSKL TVDKSRWQQG NVFSCSVMHE ALHNHYTQKS LSLSPGK [SEQ ID NO: 28]

Exendin-4-IgG1 amino acid sequence is represented by SEQ ID
25 NO: 29.

1 HGEGTFTSDL SKQMEEEAVR LFIEWLKNGG PSSGAPPPSA EPKSCDKTHT CPPCPAPELL
61 GGPSVFLFPP KPKDTLMISR TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREGQ
121 YNSTYRVVSV LTQLHQDWLN GKEYKCKVSN KALPAPIEKT ISKAKGQPREGQVYTLPPSR
181 EEMTKNQVSL TCLVKGFYPS DIAVEWESNG QPENNYKTTP PVLDSDGSFF LYSKLTVDKS
30 241 RWQQGNVFSC SVMHEALHNH YTQKSLSLSP GK [SEQ ID NO: 29]

- 85 -

Exendin-4-C2-IgG1 amino acid sequence is represented by SEQ ID NO: 30.

1 HGETGFTSDL SKQMEEEAVR LFIEWLKNGG PSSGASSGAA EPKSCDKTHT CPPCPAPELL
 5 61 GGPSVFLFPP KPKDTLMISR TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ
 121 YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN KALPAPIEKT ISKAKGQPREG QPVYTLPPSR
 181 EEMTKNQVSL TCLVKGFYPS DIAVEWESNG QPVENYKTPP PVLDSDGSFF LYSKLTVDKS
 241 RWQQGNVFSC SVMHEALHNH YTQKSLSLSP GK [SEQ ID NO: 30]

10 Exendin-4-Linker-IgG1 amino acid sequence is represent by SEQ ID NO: 31.

1 HGETGFTSDL SKQMEEEAVR LFIEWLKNGG PSSGAPPPSG GGGSGGGGSG GGGSAEPKSC
 61 DKTHTCPPCP APELLGGPSV FLFPPKPKDT LMISRTPEVT CVVVDVSHED PEVKFNWYVD
 121 GVEVHNNAKTK PREEQYNSTY RVVSVLTVLH QDWLNGKEYK CKVSNKALPA PIEKTISKAK
 15 181 GQPREPQVYT LPPSREEMTK NQVSLTCLVK GFYPSDIAVE WESNGQPEENN YKTTPPVLD
 241 DGSFFFLYSKL TVDKSRWQQG NVFSCSVMHE ALHNHYTQKS LSLSPGK [SEQ ID NO: 31]

Table 5a: *In vitro* GLP-1 analog activity

GLP-1 Receptor

	<u>GLP-1 Compound</u>	<u>Activation</u>
20	GLP-1(7-37)OH	1.0
	Val ⁸ -GLP-1(7-37)OH	0.47 (n = 6)
25	Gly ⁸ -His ¹¹ -GLP-1(7-37)OH	0.282
	Val ⁸ -Ala ¹¹ -GLP-1(7-37)OH	0.021
30	Val ⁸ -Lys ¹¹ -GLP-1(7-37)OH	0.001
	Val ⁸ -Tyr ¹² -GLP-1(7-37)OH	0.81
	Val ⁸ -Glu ¹⁶ -GLP-1(7-37)OH	0.047

- 86 -

	Val ⁸ -Ala ¹⁶ -GLP-1(7-37)OH	0.112
5	Val ⁸ -Tyr ¹⁶ -GLP-1(7-37)OH	1.175
	Val ⁸ -Lys ²⁰ -GLP-1(7-37)OH	0.33
	Gln ²² -GLP-1(7-37)OH	0.42
10	Val ⁸ -Ala ²² -GLP-1(7-37)OH	0.56
	Val ⁸ -Ser ²² -GLP-1(7-37)OH	0.50
	Val ⁸ -Asp ²² -GLP-1(7-37)OH	0.40
15	Val ⁸ -Glu ²² -GLP-1(7-37)OH	1.29
	Val ⁸ -Lys ²² -GLP-1(7-37)OH	0.58
20	Val ⁸ -Pro ²² -GLP-1(7-37)OH	0.01
	Val ⁸ -His ²² -GLP-1(7-37)OH	0.14
	Val ⁸ -Lys ²² -GLP-1(7-36)NH ₂	0.53
25	Val ⁸ -Glu ²² -GLP-1(7-36)NH ₂	1.0
	Gly ⁸ -Glu ²² -GLP-1(7-37)OH	1.07
30	Val ⁸ -Lys ²³ -GLP-1(7-37)OH	0.18
	Val ⁸ -His ²⁴ -GLP-1(7-37)OH	0.007

- 87 -

	Val ⁸ -Lys ²⁴ -GLP-1 (7-37) OH	0.02
	Val ⁸ -His ²⁶ -GLP-1 (7-37) OH	1.6
5	Val ⁸ -Glu ²⁶ -GLP-1 (7-37) OH	1.5
	Val ⁸ -His ²⁷ -GLP-1 (7-37) OH	0.37
	Val ⁸ -Ala ²⁷ -GLP-1 (7-37) OH	0.47
10	Gly ⁸ -Glu ³⁰ -GLP-1 (7-37) OH	0.29
	Val ⁸ -Glu ³⁰ -GLP-1 (7-37) OH	0.29
15	Val ⁸ -Asp ³⁰ -GLP-1 (7-37) OH	0.15
	Val ⁸ -Ser ³⁰ -GLP-1 (7-37) OH	0.19
	Val ⁸ -His ³⁰ -GLP-1 (7-37) OH	0.19
20	Val ⁸ -Glu ³³ -GLP-1 (7-37) OH	0.039
	Val ⁸ -Ala ³³ -GLP-1 (7-37) OH	0.1
25	Val ⁸ -Gly ³³ -GLP-1 (7-37) OH	0.01
	Val ⁸ -Glu ³⁴ -GLP-1 (7-37) OH	0.17
	Val ⁸ -Pro ³⁵ -GLP-1 (7-37) OH	0.094
30	Val ⁸ -His ³⁵ -GLP-1 (7-37) OH	0.41
	Val ⁸ -Glu ³⁵ -GLP-1 (7-37) OH	0.15

- 88 -

	Val ⁸ -Glu ³⁶ -GLP-1 (7-37) OH	0.11
5	Val ⁸ -His ³⁶ -GLP-1 (7-37) OH	0.22
	Val ⁸ -His ³⁷ -GLP-1 (7-37) OH	0.33
10	Val ⁸ -Leu ¹⁶ -Glu ²⁶ -GLP-1 (7-37) OH	0.23
	Val ⁸ -Lys ²² -Glu ³⁰ -GLP-1 (7-37) OH	0.37
	Val ⁸ -Lys ²² -Glu ²³ -GLP-1 (7-37) OH	0.35
15	Val ⁸ -Glu ²² -Ala ²⁷ -GLP-1 (7-37) OH	1.02
	Val ⁸ -Glu ²² -Lys ²³ -GLP-1 (7-37) OH	1.43
	Val ⁸ -Lys ³³ -Val ³⁴ -GLP-1 (7-37) OH	0.08
20	Val ⁸ -Lys ³³ -Asn ³⁴ -GLP-1 (7-37) OH	0.09
	Val ⁸ -Gly ³⁴ -Lys ³⁵ -GLP-1 (7-37) OH	0.34
25	Val ⁸ -Gly ³⁶ -Pro ³⁷ -GLP-1 (7-37) NH ₂	0.53

- 89 -

Table 5b. *In vitro* GLP-1 analog activity

GLP-1 Compound	GLP-1 Receptor Activation
GLP-1 (7-37) OH	1.0
Val ⁸ -GLP-1 (7-37) OH	0.47
Gly ⁸ -GLP-1 (7-37) OH	0.80
Val ⁸ -Tyr ¹² -GLP-1 (7-37) OH	0.80
Val ⁸ -Tyr ¹² -GLP-1 (7-36) NH ₂	0.52
Val ⁸ -Trp ¹² -GLP-1 (7-37) OH	0.52
Val ⁸ -Leu ¹⁶ -GLP-1 (7-37) OH	0.52
Val ⁸ -Val ¹⁶ -GLP-1 (7-37) OH	0.52
Val ⁸ -Tyr ¹⁶ -GLP-1 (7-37) OH	1.18
Gly ⁸ -Glu ²² -GLP-1 (7-37) OH	1.03
Val ⁸ -Leu ²⁵ -GLP-1 (7-37) OH	0.24
Val ⁸ -Tyr ¹² -Tyr ¹⁶ -GLP-1 (7-37) OH	0.70
Val ⁸ -Trp ¹² -Glu ²² -GLP-1 (7-37) OH	0.80
Val ⁸ -Tyr ¹² -Glu ²² -GLP-1 (7-37) OH	1.27
Val ⁸ -Tyr ¹⁶ -Phe ¹⁹ -GLP-1 (7-37) OH	1.32
Val ⁸ -Tyr ¹⁶ -Glu ²² -GLP-1 (7-37) OH	1.69, 1.79
Val ⁸ -Trp ¹⁶ -Glu ²² -GLP-1 (7-37) OH	2.30, 2.16
Val ⁸ -Leu ¹⁶ -Glu ²² -GLP-1 (7-37) OH	2.02
Val ⁸ -Ile ¹⁶ -Glu ²² -GLP-1 (7-37) OH	1.55
Val ⁸ -Phe ¹⁶ -Glu ²² -GLP-1 (7-37) OH	1.08
Val ⁸ -Trp ¹⁸ -Glu ²² -GLP-1 (7-37) OH	1.50, 3.10
Val ⁸ -Tyr ¹⁸ -Glu ²² -GLP-1 (7-37) OH	2.40, 2.77
Val ⁸ -Phe ¹⁸ -Glu ²² -GLP-1 (7-37) OH	0.94
Val ⁸ -Ile ¹⁸ -Glu ²² -GLP-1 (7-37) OH	1.88
Val ⁸ -Lys ¹⁸ -Glu ²² -GLP-1 (7-37) OH	1.18
Val ⁸ -Trp ¹⁹ -Glu ²² -GLP-1 (7-37) OH	1.50
Val ⁸ -Phe ¹⁹ -Glu ²² -GLP-1 (7-37) OH	0.70
Val ⁸ -Phe ²⁰ -Glu ²² -GLP-1 (7-37) OH	1.27

- 90 -

Val ⁸ -Glu ²² -Leu ²⁵ -GLP-1 (7-37) OH	1.32
Val ⁸ -Glu ²² -Ile ²⁵ -GLP-1 (7-37) OH	1.46
Val ⁸ -Glu ²² -Val ²⁵ -GLP-1 (7-37) OH	2.21, 1.36
Val ⁸ -Glu ²² -Ile ²⁷ -GLP-1 (7-37) OH	0.94
Val ⁸ -Glu ²² -Ala ²⁷ -GLP-1 (7-37) OH	1.03
Val ⁸ -Glu ²² -Ile ³³ -GLP-1 (7-37) OH	2.21, 1.79, 1.60
Val ⁸ -Asp ⁹ -Ile ¹¹ -Tyr ¹⁶ -Glu ²² -GLP-1 (7-37) OH	2.02
Val ⁸ -Tyr ¹⁶ -Trp ¹⁹ -Glu ²² -GLP-1 (7-37) OH	1.64
Val ⁸ -Trp ¹⁶ -Glu ²² -Val ²⁵ -Ile ³³ -GLP-1 (7-37) OH	2.35
Val ⁸ -Trp ¹⁶ -Glu ²² -Ile ³³ -GLP-1 (7-37) OH	1.93
Val ⁸ -Glu ²² -Val ²⁵ -Ile ³³ -GLP-1 (7-37) OH	2.30, 2.73, 3.15
Val ⁸ -Trp ¹⁶ -Glu ²² -Val ²⁵ -GLP-1 (7-37) OH	2.07
Val ⁸ -Cys ¹⁶ -Lys ²⁶ -GLP-1 (7-37) OH	1.97
Val ⁸ -Cys ¹⁶ -Lys ²⁶ -Arg ³⁴ -GLP-1 (7-37) OH	2.4, 1.9

Table 6: In vitro activity of GLP/Exendin analogs

Peptide Sequence	In Vitro Activity (% of Val ⁸ -GLP-1 (7-37) OH)
HGETFTSDL S SKQMEEEAVRLFIEWLKNGGP-NH2	6.21
HGETFTSDL S SKQMEEEAVRLFIEWLKNGGPSSGAPP PS-NH2	6.75, 3.25
HVEGTFTSDL S SKQMEEEAVRLFIAWLVKGRG	2.86
HVEGTFTSDVSSYLEEEAVRLFIAWLVKGRG	1.47
HVEGTFTSDL S SKQMEGQAAKEFIAWLVKGRG	0.11

- 91 -

HVEGTFTSDVSKQMEGQAAKEFIAWLVKGRG	0.04
HGEGTFTSDL SKQMEGQAAKEFIEWLKNGGP-NH2	1.44
HGEGTFTSDL SKQMEEEAAKEFIEWLKNGGP-NH2	2.80
HGEGTFTSDVSSYLEEEEAVRLFIEWLKNGGP-NH2	5.40
HGEGTFTSDLSSYLEEEEAVRLFIEWLKNGGP-NH2	5.07
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRPSSGAPPPS -NH2	3.30
HAEGTFTSDVSKQLEEEEAAKEFIAWLVKGRG	2.15
HVEGTFTSDVSSYLEGQAAKEFIEWLKNGGP-NH2	2.36
HGEGTFTSDL SKQMEEEAVRLFIAWLVKGRG	3.25
HVEGTFTSDVSSYLEEEEAAKEFIAWLVKGRG	1.00
HVEGTFTSDVSSYLEGQAAKEFIAWLKNGRG	0.20
HVEGTFTSDVSSYLEGQAAKEFIAWLVKGRG	1.00
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG	2.12

Example 7: In vivo pharmacokinetics of Val⁸-GLP-1-IgG1
and Val⁸-GLP-1-HSA:

A pharmacokinetic study of Val⁸-GLP-1-IgG1 and Val⁸-

5 GLP-1-HSA was performed in cynomolgus monkeys. Monkeys were dosed at 5.6 nmoles/kg with either purified Val⁸-GLP-1-IgG1 or Val⁸-GLP-1-HSA. The compounds were administered as an intravenous bolus administration. Blood was collected pre-dose and at 0.083, 0.25, 0.5, 1, 4, 8, 12, 24, 48, 72, 96, 120, 144, 168, and 216 hours post-dose into tubes containing EDTA. Plasma concentrations of immunoreactive Val⁸-GLP-1 were determined using a radioimmunoassay that utilizes a polyclonal antiserum whose primary specificity is for the N-terminal (7-16) region of Val⁸-GLP-1(7-37).

10 15 Figure 9 depicts the plasma concentration of Val⁸-GLP-1-Fc and Val⁸-GLP-1-Linker-HSA following a single intravenous dose to two cynomolgus monkeys. The Fc fusion protein had a half-life of approximately 45 hours and the albumin fusion had a half-life of approximately 87 hours.

- 92 -

Example 8: In vivo pharmacodynamics of Exendin-4-IgG1:

Two chronically cannulated normal male beagle dogs were studied after an overnight fast. Arterial and venous vascular access ports were accessed, and a catheter was 5 inserted percutaneously into a cephalic vein and secured. Animals were placed in cages, and their catheters were attached to a swivel/tether system. A solution containing the fusion protein Exendin-4-IgG1 (11.8 μ M) was injected intravenously (1.0 nmol/kg) through the cephalic vein 10 catheter. The catheter was then cleared with 10 ml of saline. Two hours later, a hyperglycemic (150 mg/dl) clamp was initiated and continued for three hours. Arterial blood samples were drawn throughout this 5-hour period for determination of plasma concentrations of the fusion 15 protein, glucose, and insulin.

The results of this study were compared to those from a similar, previous study in which both of the animals had received a bolus of saline, s.c., and three hours later were studied using a 3-hour hyperglycemic (150 mg/dl) clamp.

20 In both sets of studies, plasma glucose concentrations were determined using a Beckman glucose analyzer. Plasma insulin concentrations were determined by employees of Linco Research, Inc. using an RIA kit developed in their laboratories. The data is illustrated in Figures 10 and 11.

25

Example 9: In vivo pharmacokinetics of
Gly⁸-Glu²²-GLP-1-CEx-Linker-IgG1:

Two groups of three normal male beagle dogs received 0.1 mg/kg of Gly⁸-Glu²²-GLP-1-CEx-Linker-IgG1 by subcutaneous 30 (SC) or intravenous (IV) administration. Plasma concentrations of Gly⁸-Glu²²-GLP-1-CEx-Linker-IgG1 immunoreactivity were determined by radioimmunoassay in samples collected from 30 minutes predose to 216 hours postdose for both the IV and SC groups. These 35 concentrations were subsequently used to determine the

- 93 -

reported pharmacokinetic parameters. The mean elimination half-life of IV administered Gly⁸-Glu²²-GLP-1-CEx-Linker-IgG1 was approximately 55 hours and the total body clearance was 1.5 mL/h/kg. The mean elimination half-life of SC 5 administered Gly⁸-Glu²²-GLP-1-CEx-Linker-IgG1 was approximately 38 hours.

WE CLAIM:

1. A heterologous fusion protein comprising a first polypeptide with a N-terminus and a C-terminus fused to a second polypeptide with a N-terminus and a C-terminus wherein the first polypeptide is a GLP-1 compound and the second polypeptide is selected from the group consisting of
 - a) human albumin;
 - b) human albumin analogs; and
 - c) fragments of human albumin,and wherein the C-terminus of the first polypeptide is fused to the N-terminus of the second polypeptide.
- 15 2. A heterologous fusion protein comprising a first polypeptide with a N-terminus and a C-terminus fused to a second polypeptide with a N-terminus and a C-terminus wherein the first polypeptide is a GLP-1 compound and the second polypeptide is selected from the group consisting of
 - a) human albumin;
 - b) human albumin analogs; and
 - c) fragments of human albumin,and wherein the C-terminus of the first polypeptide is fused to the N-terminus of the second polypeptide via a peptide linker.
- 25 3. The heterologous fusion protein of the Claim 2 wherein the peptide linker is selected from the group consisting of:
 - a) a glycine rich peptide;
 - b) a peptide having the sequence $[\text{Gly-Gly-Gly-Gly-Ser}]_n$ where n is 1, 2, 3, 4, 5 or 6; and
 - c) a peptide having the sequence $[\text{Gly-Gly-Gly-Gly-Ser}]_3$.

- 95 -

4. The heterologous fusion protein of Claims 1, 2, or 3 wherein the GLP-1 compound comprises the sequence of formula 1 [SEQ ID NO: 2]

5 7 8 9 10 11 12 13 14 15 16 17
His-Xaa-Xaa-Gly-Xaa-Phe-Thr-Xaa-Asp-Xaa-Xaa-
18 19 20 21 22 23 24 25 26 27 28
Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Phe-
29 30 31 32 33 34 35 36 37 38 39
10 Ile-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-
40 41 42 43 44 45
Xaa-Xaa-Xaa-Xaa-Xaa-Xaa

Formula I (SEQ ID NO: 2)

15 wherein:

Xaa at position 8 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;
Xaa at position 9 is Glu, Asp, or Lys;
Xaa at position 11 is Thr, Ala, Gly, Ser, Leu, Ile,
20 Val, Glu, Asp, or Lys;
Xaa at position 14 is Ser, Ala, Gly, Thr, Leu, Ile,
Val, Glu, Asp, or Lys;
Xaa at position 16 is Val, Ala, Gly, Ser, Thr, Leu,
Ile, Tyr, Glu, Asp, Trp, or Lys;
25 Xaa at position 17 is Ser, Ala, Gly, Thr, Leu, Ile,
Val, Glu, Asp, or Lys;
Xaa at position 18 is Ser, Ala, Gly, Thr, Leu, Ile,
Val, Glu, Asp, Trp, Tyr, or Lys;
Xaa at position 19 is Tyr, Phe, Trp, Glu, Asp, Gln, or
30 Lys;
Xaa at position 20 is Leu, Ala, Gly, Ser, Thr, Ile,
Val, Glu, Asp, Met, Trp, Tyr, or Lys;
Xaa at position 21 is Glu, Asp, or Lys;
Xaa at position 22 is Gly, Ala, Ser, Thr, Leu, Ile,
35 Val, Glu, Asp, or Lys;

- 96 -

Xaa at position 23 is Gln, Asn, Arg, Glu, Asp, or Lys;
Xaa at position 24 is Ala, Gly, Ser, Thr, Leu, Ile,
Val, Arg, Glu, Asp, or Lys;
Xaa at position 25 is Ala, Gly, Ser, Thr, Leu, Ile,
5 Val, Glu, Asp, or Lys;
Xaa at position 26 is Lys, Arg, Gln, Glu, Asp, or His;
Xaa at position 27 is Leu, Glu, Asp, or Lys;
Xaa at position 30 is Ala, Gly, Ser, Thr, Leu, Ile,
Val, Glu, Asp, or Lys;
10 Xaa at position 31 is Trp, Phe, Tyr, Glu, Asp, or Lys;
Xaa at position 32 is Leu, Gly, Ala, Ser, Thr, Ile,
Val, Glu, Asp, or Lys;
Xaa at position 33 is Val, Gly, Ala, Ser, Thr, Leu,
Ile, Glu, Asp, or Lys;
15 Xaa at position 34 is Asn, Lys, Arg, Glu, Asp, or His;
Xaa at position 35 is Gly, Ala, Ser, Thr, Leu, Ile,
Val, Glu, Asp, or Lys;
Xaa at position 36 is Gly, Arg, Lys, Glu, Asp, or His;
Xaa at position 37 is Pro, Gly, Ala, Ser, Thr, Leu,
20 Ile, Val, Glu, Asp, or Lys, or is deleted;
Xaa at position 38 is Ser, Arg, Lys, Glu, Asp, or His,
or is deleted;
Xaa at position 39 is Ser, Arg, Lys, Glu, Asp, or His,
or is deleted;
25 Xaa at position 40 is Gly, Asp, Glu, or Lys, or is
deleted;
Xaa at position 41 is Ala, Phe, Trp, Tyr, Glu, Asp, or
Lys, or is deleted;
Xaa at position 42 is Ser, Pro, Lys, Glu, or Asp, or is
30 deleted;
Xaa at position 43 is Ser, Pro, Glu, Asp, or Lys, or is
deleted;
Xaa at position 44 is Gly, Pro, Glu, Asp, or Lys, or is
deleted; and

- 97 -

Xaa at position 45 is Ala, Ser, Val, Glu, Asp, or Lys, or is

deleted;

provided that when the amino acid at position 37, 38, 39, 5 40, 41, 42, 43, or 44 is deleted, then each amino acid downstream of that amino acid is also deleted.

5. The heterologous fusion protein of Claims 1, 2, or 3 wherein the GLP-1 compound comprises the sequence of formula 10 II (SEQ ID NO: 3):

7	8	9	10	11	12	13	14	15	16	17
Xaa-Xaa-Xaa-Gly-Xaa-Xaa-Thr-Ser-Asp-Xaa-Ser-										
18	19	20	21	22	23	24	25	26	27	28
15	Xaa-Tyr-Leu-Glu-Xaa-Xaa-Xaa-Ala-Xaa-Xaa-Phe-									
29	30	31	32	33	34	35	36	37		
Ile-Xaa-Xaa-Leu-Xaa-Xaa-Xaa-Xaa-Xaa										

Formula II (SEQ ID NO: 3)

20 wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

25 Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

Xaa at position 9 is: Thr, Ser, Arg, Lys, Trp, Phe, Tyr, Glu, or His;

Xaa at position 11 is: Asp, Glu, Arg, Thr, Ala, Lys, or His;

30 Xaa at position 12 is: His, Trp, Phe, or Tyr;

Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Tyr, Glu, or Ala;

Xaa at position 18 is: His, Pro, Asp, Glu, Arg, Ser, Ala, or Lys;

35 Xaa at position 19 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or

- 98 -

Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, Gln, or Arg;
Xaa at position 24 is: Glu, Arg, Ala, or Lys;
Xaa at position 26 is: Trp, Tyr, Phe, Asp, Lys, Glu, or His;
5 Xaa at position 27 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, or
Lys;
Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;
Xaa at position 31 is: Asp, Glu, Ser, Thr, Arg, Trp, or Lys;
Xaa at position 33 is: Asp, Arg, Val, Lys, Ala, Gly, or Glu;
10 Xaa at position 34 is: Glu, Lys, or Asp;
Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe,
Asp, Gly, Pro, His, or Glu;
Xaa at position 36 is: Thr, Ser, Asp, Trp, Tyr, Phe, Arg,
Glu, or His;
15 Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp,
Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

6. The heterologous fusion protein of Claims 1, 2, or 3
wherein the GLP-1 compound comprises the sequence of formula
20 III (SEQ ID NO: 4):

7	8	9	10	11	12	13	14	15	16	17
Xaa-Xaa-Glu-Gly-Xaa-Xaa-Thr-Ser-Asp-Xaa-Ser-										
18	19	20	21	22	23	24	25	26	27	28
Ser-Tyr-Leu-Glu-Xaa-Xaa-Xaa-Ala-Xaa-Xaa-Phe-										
25	29	30	31	32	33	34	35	36	37	
Ile-Ala-Xaa-Leu-Xaa-Xaa-Xaa-Xaa-Xaa										
formula III (SEQ ID NO: 4)										

wherein:

30 Xaa at position 7 is: L-histidine, D-histidine, desamino-
histidine, 2-amino-histidine, β -hydroxy-histidine,
homohistidine, α -fluoromethyl-histidine or α -methyl-
histidine;
Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;
35 Xaa at position 11 is: Asp, Glu, Arg, Thr, Ala, Lys, or His;

- 99 -

Xaa at position 12 is: His, Trp, Phe, or Tyr;

Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Glu, or Ala;

Xaa at position 22: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or 5 Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

Xaa at position 24 is: Glu, His, Ala, or Lys;

Xaa at position 25 is: Asp, Lys, Glu, or His;

Xaa at position 27 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, 10 or Lys;

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

Xaa at position 33 is: Asp, Arg, Val, Lys, Ala, Gly, or 15 Glu;

Xaa at position 34 is: Glu, Lys, or Asp;

Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe, 20 Asp, Gly, Pro, His, or Glu;

Xaa at position 36 is: Arg, Glu, or His;

Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, 25 Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

20

7. The heterologous fusion protein of Claim 1, 2, or 3 wherein the GLP-1 compound comprises the sequence of formula IV (SEQ ID NO: 5):

7 8 9 10 11 12 13 14 15 16 17
25 Xaa-Xaa-Glu-Gly-Thr-Xaa-Thr-Ser-Asp-Xaa-Ser-
18 19 20 21 22 23 24 25 26 27 28
26 Ser-Tyr-Leu-Glu-Xaa-Xaa-Ala-Ala-Xaa-Glu-Phe-
27 28 29 30 31 32 33 34 35 36 37
29 Ile-Xaa-Trp-Leu-Val-Lys-Xaa-Arg-Xaa
30 formula IV (SEQ ID NO: 5)

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine,

- 100 -

homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

Xaa at position 12 is: His, Trp, Phe, or Tyr;

5 Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Glu, or Ala;

Xaa at position 22 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

10 Xaa at position 26 is: Asp, Lys, Glu, or His;

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe, Asp, Gly, Pro, His, or Glu;

Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp,

15 Tyr, Phe, His, -NH_2 , Gly, Gly-Pro, or Gly-Pro- NH_2 , or is deleted.

8. The heterologous fusion protein of Claims 1, 2, or 3 wherein the GLP-1 compound comprises the sequence of formula

20 V (SEQ ID NO: 6)

7 8 9 10 11 12 13 14 15 16 17
Xaa-Xaa-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val-Ser-
18 19 20 21 22 23 24 25 26 27 28
25 Ser-Tyr-Leu-Glu-Xaa-Xaa-Ala-Ala-Lys-Xaa-Phe-
29 30 31 32 33 34 35 36 37
Ile-Xaa-Trp-Leu-Val-Lys-Gly-Arg-Xaa
formula V (SEQ ID NO: 6)

30 wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

35 Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

- 101 -

Xaa at position 22 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

Xaa at position 24 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, or Lys

5 or Lys

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

10 9. The heterologous fusion protein of Claims 1, 2, or 3
wherein the GLP-1 compound comprises the sequence of formula
VIII (SEQ ID NO:11).

7 8 9 10 11 12 13 14 15 16 17
 15 Xaa-Xaa-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Xaa-Ser-
 18 19 20 21 22 23 24 25 26 27 28
 29 30 31 32 33 34 35 36 37
 20 Xaa-Xaa-Xaa-Glu-Xaa-Xaa-Ala-Xaa-Xaa-Xaa-Phe-
 38 39 40 41 42 43 44 45 46 47
 48 49 50 51 52 53 54 55 56 57
 58 59 60 61 62 63 64 65 66 67
 68 69 70 71 72 73 74 75 76 77
 78 79 80 81 82 83 84 85 86 87
 88 89 90 91 92 93 94 95 96 97
 98 99 100 101 102 103 104 105 106 107
 108 109 110 111 112 113 114 115 116 117
 118 119 120 121 122 123 124 125 126 127
 128 129 130 131 132 133 134 135 136 137
 138 139 140 141 142 143 144 145 146 147
 148 149 150 151 152 153 154 155 156 157
 158 159 160 161 162 163 164 165 166 167
 168 169 170 171 172 173 174 175 176 177
 178 179 180 181 182 183 184 185 186 187
 188 189 190 191 192 193 194 195 196 197
 198 199 200 201 202 203 204 205 206 207
 208 209 210 211 212 213 214 215 216 217
 218 219 220 221 222 223 224 225 226 227
 228 229 230 231 232 233 234 235 236 237
 238 239 240 241 242 243 244 245 246 247
 248 249 250 251 252 253 254 255 256 257
 258 259 260 261 262 263 264 265 266 267

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine,

homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

Xaa at position 8 is: Gly, Ala, or Val;

Xaa at position 16 is: Leu or Val;

Xaa at position 18 is Lys or Ser;

30 Xaa at position 19 is: Gln or Tyr;

Xaa at position 20 is: Met or Leu;

Xaa at position 22 is: Glu or Gln;

Xaa at position 23 is: Glu, or Gln;

Xaa at position 25 is: Val or Ala

35 Xaa at position 26 is: Arg or Lys;

- 102 -

Xaa at position 27 is Leu or Glu;
Xaa at position 30 is: Glu or Ala;
Xaa at position 33 is: Val or Lys;
Xaa at position 34 is: Asn or Lys;
5 Xaa at position 36 is: Gly or Arg; and
Xaa at position 37 is: Gly, Pro, Pro-Ser-Ser-Gly-Ala-Pro-
Pro-Pro-Ser, or is absent.

10. The heterologous fusion protein of claims 1 through 9
10 wherein the GLP-1 compound has no more than 6 amino acids
that are different from the corresponding amino acid in GLP-
1(7-37)OH, GLP-1(7-36)OH, or Exendin-4.

11. The heterologous fusion protein of Claim 10 wherein the
15 GLP-1 compound has no more than 5 amino acids that differ
from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-
36)OH, or Exendin-4.

12. The heterologous fusion protein of Claim 11 wherein the
20 GLP-1 compound has no more than 4 amino acids that differ
from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-
36)OH, or Exendin-4.

13. The heterologous fusion protein of Claim 12 wherein the
25 GLP-1 compound has no more than 3 amino acids that differ
from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-
36)OH, or Exendin-4.

14. The heterologous fusion protein of Claim 13 wherein the
30 GLP-1 compound has no more than 2 amino acids that differ
from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-
36)OH, or Exendin-4.

15. The heterologous fusion protein of any one of Claims 1
35 through 14 wherein Xaa at position 8 is glycine or valine.

16. The heterologous fusion protein of Claim 8 wherein the GLP-1 compound has no more than 2 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH or GLP-
5 1(7-36)OH and Xaa at position 8 is glycine or valine and Xaa at position 30 is alanine, glutamic acid, aspartic acid, serine, or histidine.

17. The heterologous fusion protein of Claim 16 wherein Xaa
10 at position 30 is Glutamic acid.

18. The heterologous fusion protein of Claim 8 wherein the GLP-1 compound has no more than 2 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH or GLP-
15 1(7-36)OH and Xaa at position 8 is glycine or valine and Xaa at position 37 is histidine, phenylalanine, tyrosine, tryptophan, aspartic acid, glutamic acid, serine, threonine, arginine, or lysine.

20 19. The heterologous fusion protein of Claim 18 wherein Xaa at position 37 is histidine.

20. The heterologous fusion protein of Claim 8 wherein the GLP-1 compound has no more than 2 amino acids that differ
25 from the corresponding amino acid in GLP-1(7-37)OH or GLP-1(7-36)OH and Xaa at position 8 is glycine, valine, leucine, isoleucine, serine, threonine, or methionine and Xaa at position 22 is aspartic acid, glutamic acid, lysine, arginine, asparagine, glutamine or histidine.

30 21. The heterologous fusion protein of Claim 20 wherein Xaa at position 22 is lysine or glutamic acid.

22. The heterologous fusion protein of Claim 8 wherein the
35 GLP-1 compound is Val⁸-GLP-1(7-37).

- 104 -

23. The heterologous fusion protein of Claim 8 wherein the GLP-1 compound is Gly⁸-GLP-1(7-37).

5 24. The heterologous fusion protein of any one of Claims 1 through 23 wherein the second polypeptide is human albumin.

25. The heterologous fusion protein of Claim 24 wherein the second polypeptide has the sequence of SEQ ID NO: 34.

10

26. The heterologous fusion protein of any one of Claims 1 through 23 wherein the second polypeptide is an N-terminal fragment of albumin.

15

27. A heterologous fusion protein comprising a first polypeptide with a N-terminus and a C-terminus fused to a second polypeptide with a N-terminus and a C-terminus wherein the first polypeptide is a GLP-1 compound and the second polypeptide is selected from the group consisting of

20

a) the Fc portion of an immunoglobulin;

b) an analog of the Fc portion of an immunoglobulin;

and

c) fragments of the Fc portion of an immunoglobulin,

and wherein the C-terminus of the first polypeptide is fused to the N-terminus of the second polypeptide.

25

28. A heterologous fusion protein comprising a first polypeptide with a N-terminus and a C-terminus fused to a second polypeptide with a N-terminus and a C-terminus

30

wherein the first polypeptide is a GLP-1 compound and the second polypeptide is selected from the group consisting of

a) the Fc portion of an immunoglobulin;

b) an analog of the Fc portion of an immunoglobulin;

and

35

5 c) fragments of the Fc portion of an immunoglobulin, and wherein the C-terminus of the first polypeptide is fused to the N-terminus of the second polypeptide via a peptide linker.

29. The heterologous fusion protein of the Claim 28 wherein the peptide linker is selected from the group consisting of:
10 a) a glycine rich peptide;
b) a peptide having the sequence [Gly-Gly-Gly-Gly-Ser]_n where n is 1, 2, 3, 4, 5, or 6; and
c) a peptide having the sequence [Gly-Gly-Gly-Gly-Ser]₃.

30. The heterologous fusion protein of any one of Claims
15 27, 28, or 29 wherein the GLP-1 compound comprises the sequence of formula 1 [SEQ ID NO: 2].

7 8 9 10 11 12 13 14 15 16 17
His-Xaa-Xaa-Gly-Xaa-Phe-Thr-Xaa-Asp-Xaa-Xaa-
20 18 19 20 21 22 23 24 25 26 27 28
Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Phe-
29 30 31 32 33 34 35 36 37 38 39
Ile-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-
40 41 42 43 44 45
25 Xaa-Xaa-Xaa-Xaa-Xaa-Xaa
Formula I (SEQ ID NO: 2)

wherein:

30 Xaa at position 8 is Ala, Gly, Ser, Thr, Leu, Ile, Val, Glu, Asp, or Lys;
Xaa at position 9 is Glu, Asp, or Lys;
Xaa at position 11 is Thr, Ala, Gly, Ser, Leu, Ile, Val, Glu, Asp, or Lys;
Xaa at position 14 is Ser, Ala, Gly, Thr, Leu, Ile, Val, Glu, Asp, or Lys;

- 106 -

Xaa at position 16 is Val, Ala, Gly, Ser, Thr, Leu, Ile,
Tyr, Glu, Asp, Trp, or Lys;

Xaa at position 17 is Ser, Ala, Gly, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;

5 Xaa at position 18 is Ser, Ala, Gly, Thr, Leu, Ile, Val,
Glu, Asp, Trp, Tyr, or Lys;

Xaa at position 19 is Tyr, Phe, Trp, Glu, Asp, Gln, or Lys;

Xaa at position 20 is Leu, Ala, Gly, Ser, Thr, Ile, Val,
Glu, Asp, Met, Trp, Tyr, or Lys;

10 Xaa at position 21 is Glu, Asp, or Lys;

Xaa at position 22 is Gly, Ala, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;

Xaa at position 23 is Gln, Asn, Arg, Glu, Asp, or Lys;

Xaa at position 24 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
15 Arg, Glu, Asp, or Lys;

Xaa at position 25 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;

Xaa at position 26 is Lys, Arg, Gln, Glu, Asp, or His;

Xaa at position 27 is Leu, Glu, Asp, or Lys;

20 Xaa at position 30 is Ala, Gly, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;

Xaa at position 31 is Trp, Phe, Tyr, Glu, Asp, or Lys;

Xaa at position 32 is Leu, Gly, Ala, Ser, Thr, Ile, Val,
Glu, Asp, or Lys;

25 Xaa at position 33 is Val, Gly, Ala, Ser, Thr, Leu, Ile,
Glu, Asp, or Lys;

Xaa at position 34 is Asn, Lys, Arg, Glu, Asp, or His;

Xaa at position 35 is Gly, Ala, Ser, Thr, Leu, Ile, Val,
Glu, Asp, or Lys;

30 Xaa at position 36 is Gly, Arg, Lys, Glu, Asp, or His;

Xaa at position 37 is Pro, Gly, Ala, Ser, Thr, Leu, Ile,
Val, Glu, Asp, or Lys, or is deleted;

Xaa at position 38 is Ser, Arg, Lys, Glu, Asp, or His, or
is deleted;

35 Xaa at position 39 is Ser, Arg, Lys, Glu, Asp, or His, or

- 107 -

is deleted;

Xaa at position 40 is Gly, Asp, Glu, or Lys, or is deleted;
Xaa at position 41 is Ala, Phe, Trp, Tyr, Glu, Asp, or Lys,
or is deleted;

5 Xaa at position 42 is Ser, Pro, Lys, Glu, or Asp, or is
deleted;

Xaa at position 43 is Ser, Pro, Glu, Asp, or Lys, or is
deleted;

Xaa at position 44 is Gly, Pro, Glu, Asp, or Lys, or is

10 deleted;

and

Xaa at position 45 is Ala, Ser, Val, Glu, Asp, or Lys, or is
deleted;

provided that when the amino acid at position 37, 38, 39,

15 40, 41, 42, 43, or 44 is deleted, then each amino acid
downstream of that amino acid is also deleted.

31. The heterologous fusion protein of Claims 27, 28, or 29

wherein the GLP-1 compound comprises the sequence of formula

20 II (SEQ ID NO: 3):

7	8	9	10	11	12	13	14	15	16	17
Xaa-Xaa-Xaa-Gly-Xaa-Xaa-Thr-Ser-Asp-Xaa-Ser-										
18	19	20	21	22	23	24	25	26	27	28
Xaa-Tyr-Leu-Glu-Xaa-Xaa-Xaa-Ala-Xaa-Xaa-Phe-										
29	30	31	32	33	34	35	36	37		
Ile-Xaa-Xaa-Leu-Xaa-Xaa-Xaa-Xaa-Xaa										

Formula II (SEQ ID NO: 3)

30 wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-
histidine, 2-amino-histidine, β -hydroxy-histidine,
homohistidine, α -fluoromethyl-histidine or α -methyl-
histidine;

35 Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

- 108 -

Xaa at position 9 is: Thr, Ser, Arg, Lys, Trp, Phe, Tyr, Glu, or His;

Xaa at position 11 is: Asp, Glu, Arg, Thr, Ala, Lys, or His;

Xaa at position 12 is: His, Trp, Phe, or Tyr;

5 Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Tyr, Glu, or Ala;

Xaa at position 18 is: His, Pro, Asp, Glu, Arg, Ser, Ala, or Lys;

Xaa at position 19 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or 10 Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, Gln, or Arg;

Xaa at position 24 is: Glu, Arg, Ala, or Lys;

Xaa at position 26 is: Trp, Tyr, Phe, Asp, Lys, Glu, or His;

Xaa at position 27 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, or 15 Lys;

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

Xaa at position 31 is: Asp, Glu, Ser, Thr, Arg, Trp, or Lys;

Xaa at position 33 is: Asp, Arg, Val, Lys, Ala, Gly, or Glu;

Xaa at position 34 is: Glu, Lys, or Asp;

20 Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe, Asp, Gly, Pro, His, or Glu;

Xaa at position 36 is: Thr, Ser, Asp, Trp, Tyr, Phe, Arg, Glu, or His;

Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, 25 Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

32. The heterologous fusion protein of Claims 27, 28, or 29 wherein the GLP-1 compound comprises the sequence of formula III (SEQ ID NO: 4):

30 7 8 9 10 11 12 13 14 15 16 17
 Xaa-Xaa-Glu-Gly-Xaa-Xaa-Thr-Ser-Asp-Xaa-Ser-
 18 19 20 21 22 23 24 25 26 27 28
 Ser-Tyr-Leu-Glu-Xaa-Xaa-Xaa-Ala-Xaa-Xaa-Phe-
 29 30 31 32 33 34 35 36 37
35 Ile-Ala-Xaa-Leu-Xaa-Xaa-Xaa-R

- 109 -

formula III (SEQ ID NO: 4)

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-
histidine, 2-amino-histidine, β -hydroxy-histidine,
homohistidine, α -fluoromethyl-histidine or α -methyl-
histidine;

Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

Xaa at position 11 is: Asp, Glu, Arg, Thr, Ala, Lys, or His;

Xaa at position 12 is: His, Trp, Phe, or Tyr;

Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp,
Val, Glu, or Ala;

Xaa at position 22: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or
Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

Xaa at position 24 is: Glu, His, Ala, or Lys;

Xaa at position 25 is: Asp, Lys, Glu, or His;

Xaa at position 27 is: Ala, Glu, His, Phe, Tyr, Trp, Arg,
or Lys;

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

Xaa at position 33 is: Asp, Arg, Val, Lys, Ala, Gly, or
Glu;

Xaa at position 34 is: Glu, Lys, or Asp;

Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe,
Asp, Gly, Pro, His, or Glu;

Xaa at position 36 is: Arg, Glu, or His;

Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp,
Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

30 33. The heterologous fusion protein of Claim 27, 28, or 29
wherein the GLP-1 compound comprises the sequence of formula
IV (SEQ ID NO: 5):

7 8 9 10 11 12 13 14 15 16 17
Xaa-Xaa-Glu-Gly-Thr-Xaa-Thr-Ser-Asp-Xaa-Ser-
35 18 19 20 21 22 23 24 25 26 27 28

- 110 -

Ser-Tyr-Leu-Glu-Xaa-Xaa-Ala-Ala-Xaa-Glu-Phe-
29 30 31 32 33 34 35 36 37
Ile-Xaa-Trp-Leu-Val-Lys-Xaa-Arg-Xaa
formula IV (SEQ ID NO: 5)

5

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

10

Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;

Xaa at position 12 is: His, Trp, Phe, or Tyr;

Xaa at position 16 is: Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Glu, or Ala;

15

Xaa at position 22 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

Xaa at position 26 is: Asp, Lys, Glu, or His;

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

20

Xaa at position 35 is: Thr, Ser, Lys, Arg, Trp, Tyr, Phe, Asp, Gly, Pro, His, or Glu;

Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

25

34. The heterologous fusion protein of Claims 27, 28, or 29 wherein the GLP-1 compound comprises the sequence of formula V (SEQ ID NO: 6)

30

7 8 9 10 11 12 13 14 15 16 17
Xaa-Xaa-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val-Ser-
18 19 20 21 22 23 24 25 26 27 28
Ser-Tyr-Leu-Glu-Xaa-Xaa-Ala-Ala-Lys-Xaa-Phe-
29 30 31 32 33 34 35 36 37
Ile-Xaa-Trp-Leu-Val-Lys-Gly-Arg-Xaa

35

formula V (SEQ ID NO: 6)

- 111 -

wherein:

Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

5

Xaa at position 8 is: Gly, Ala, Val, Leu, Ile, Ser, or Thr;
Xaa at position 22 is: Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

10 Xaa at position 23 is: His, Asp, Lys, Glu, or Gln;

Xaa at position 24 is: Ala, Glu, His, Phe, Tyr, Trp, Arg, or Lys

Xaa at position 30 is: Ala, Glu, Asp, Ser, or His;

Xaa at position 37 is: Lys, Arg, Thr, Ser, Glu, Asp, Trp, 15 Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

35. The heterologous fusion protein of Claims 27, 28, or 29 wherein the GLP-1 compound comprises the sequence of formula VIII (SEQ ID NO:11).

20

	7	8	9	10	11	12	13	14	15	16	17
	Xaa-Xaa-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Xaa-Ser-										
	18	19	20	21	22	23	24	25	26	27	28
	Xaa-Xaa-Xaa-Glu-Xaa-Xaa-Ala-Xaa-Xaa-Xaa-Phe-										
25	29	30	31	32	33	34	35	36	37		
	Ile-Xaa-Trp-Leu-Xaa-Xaa-Gly-Xaa-R										
	formula VIII (SEQ ID NO: 11)										

wherein:

30 Xaa at position 7 is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine, homohistidine, α -fluoromethyl-histidine or α -methyl-histidine;

Xaa at position 8 is: Gly, Ala, or Val;

35 Xaa at position 16 is: Leu or Val;

- 112 -

Xaa at position 18 is Lys or Ser;
Xaa at position 19 is: Gln or Tyr;
Xaa at position 20 is: Met or Leu;
Xaa at position 22 is: Glu or Gln;
5 Xaa at position 23 is: Glu, or Gln;
Xaa at position 25 is: Val or Ala;
Xaa at position 26 is: Arg or Lys;
Xaa at position 27 is Leu or Glu;
Xaa at position 30 is: Glu or Ala;
10 Xaa at position 33 is: Val or Lys;
Xaa at position 34 is: Asn or Lys;
Xaa at position 36 is: Gly or Arg; and
Xaa at position 37 is: Gly, Pro, Pro-Ser-Ser-Gly-Ala-Pro-
Pro-Pro-Ser, or is absent.

15

36. The heterologous fusion protein of claims 27 through 35 wherein the GLP-1 compound has no more than 6 amino acids that are different from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4.

20

37. The heterologous fusion protein of Claim 36 wherein the GLP-1 compound has no more than 5 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4.

25

38. The heterologous fusion protein of Claim 37 wherein the GLP-1 compound has no more than 4 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4.

30

39. The heterologous fusion protein of Claim 38 wherein the GLP-1 compound has no more than 3 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4.

35

- 113 -

40. The heterologous fusion protein of Claim 39 wherein the GLP-1 compound has no more than 2 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH, GLP-1(7-36)OH, or Exendin-4.

5

41. The heterologous fusion protein of any one of Claims 27 through 40 wherein Xaa at position 8 is glycine or valine.

42. The heterologous fusion protein of Claim 34 wherein the GLP-1 compound has no more than 2 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH or GLP-1(7-36)OH and Xaa at position 8 is glycine or valine and Xaa at position 30 is alanine, glutamic acid, aspartic acid, serine, or histidine.

10

43. The heterologous fusion protein of Claim 42 wherein Xaa at position 30 is Glutamic acid.

15

44. The heterologous fusion protein of Claim 34 wherein the GLP-1 compound has no more than 2 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH or GLP-1(7-36)OH and Xaa at position 8 is glycine or valine and Xaa at position 37 is histidine, phenylalanine, tyrosine, tryptophan, aspartic acid, glutamic acid, serine, threonine, 20 arginine, or lysine.

25

45. The heterologous fusion protein of Claim 44 wherein Xaa at position 37 is histidine.

46. The heterologous fusion protein of Claim 34 wherein the GLP-1 compound has no more than 2 amino acids that differ from the corresponding amino acid in GLP-1(7-37)OH or GLP-
5 1(7-36)OH and Xaa at position 8 is glycine, valine, leucine, isoleucine, serine, threonine, or methionine and Xaa at position 22 is aspartic acid, glutamic acid, lysine, arginine, asparagine, glutamine or histidine.

10 47. The heterologous fusion protein of Claim 46 wherein Xaa at position 22 is lysine or glutamic acid.

48. The heterologous fusion protein of Claim 34 wherein the GLP-1 compound is Val-8-GLP-1(7-37).

15 49. The heterologous fusion protein of Claim 34 wherein the GLP-1 compound is Gly-8-GLP-1(7-37).

20 50. The heterologous fusion protein of any one of Claims 27 through 49 wherein the second polypeptide is the Fc portion of an Ig selected from the group consisting of: IgG1, IgG2, IgG3, IgG4, IgE, IgA, IgD, or IgM.

25 51. The heterologous fusion protein of any one of Claims 27 through 50 wherein the second polypeptide is the Fc portion of an Ig selected from the group consisting of: IgG1, IgG2, IgG3, and IgG4.

30 52. The heterologous fusion protein of Claim 51 wherein the second polypeptide is the Fc portion of an IgG1 immunoglobulin.

35 53. The heterologous fusion protein of Claim 51 wherein the second polypeptide is the Fc portion of an IgG4 immunoglobulin.

- 115 -

54. The heterologous fusion protein of Claims 50 through 53 wherein the IgG is human.

5 55. The heterologous fusion protein of any one of Claims 27 through 54 wherein the Fc portion comprises the hinge, CH2, and CH3 domains.

10 56. The heterologous fusion protein of Claim 52 wherein the second polypeptide has the sequence of SEQ ID NO: 32.

57. A polynucleotide encoding a heterologous fusion protein of any one of Claims 1 through 56.

15 58. A vector comprising the polynucleotide of Claim 57.

59. A host cell comprising the vector of Claim 58.

20 60. A host cell expressing at least one heterologous fusion protein of any one of Claims 1 through 56.

61. The host cell of Claim 60 wherein said host cell is a CHO cell.

25 62. A process for producing a heterologous fusion protein comprising the steps of transcribing and translating a polynucleotide of Claim 57 under conditions wherein the heterologous fusion protein is expressed in detectable amounts.

30

63. A method for normalizing blood glucose levels in a mammal in need thereof comprising the administration of a therapeutically effective amount of the heterologous fusion protein of any one of Claims 1 through 56.

35

- 116 -

64. A method of treating a patient with non-insulin dependent diabetes mellitus comprising the administration of a therapeutically effective amount of the heterologous fusion protein of any one of Claims 1 through 56.

5

65. A method of treating obesity comprising the administration of a therapeutically effective amount of the heterologous fusion protein of any one of Claims 1 through 56.

10

66. The use of a heterologous fusion protein as claimed in any one of Claims 1 through 56 for the manufacture of a medicament for the treatment of patients with non-insulin dependent diabetes mellitus.

15

67. The use of a heterologous fusion protein as claimed in any one of Claims 1 through 56 for the manufacture of a medicament for the treatment of patients with obesity.

20

68. A pharmaceutical formulation adapted for the treatment of patients with non-insulin dependent diabetes comprising a heterologous fusion protein of any one of Claims 1 through 56.

25

69. The heterologous fusion protein of Claims 1, 2, or 3 wherein the GLP-1 compound comprises the sequence of formula IX [SEQ ID NO: 12]

30

Xaa₇-Xaa₈-Glu-Gly-Thr-Xaa₁₂-Thr-Ser-Asp-Xaa₁₆-Ser-Xaa₁₈-Xaa₁₉-Xaa₂₀-Glu-Xaa₂₂-Gln-Ala-Xaa₂₅-Lys-Xaa₂₇-Phe-Ile-Xaa₃₀-Trp-Leu-Xaa₃₃-Lys-Gly-Arg-Xaa₃₇
Formula IX (SEQ ID NO: 12)

wherein:

35

Xaa₇ is: L-histidine, D-histidine, desamino-histidine, 2-

- 117 -

amino-histidine, β -hydroxy-histidine,
homohistidine, α -fluoromethyl-histidine, or α -
methyl-histidine;

Xaa₈ is: Ala, Gly, Val, Leu, Ile, Ser, or Thr;

5 Xaa₁₂ is: Phe, Trp, or Tyr;

Xaa₁₆ is: Val, Trp, Ile, Leu, Phe, or Tyr;

Xaa₁₈ is: Ser, Trp, Tyr, Phe, Lys, Ile, Leu, Val;

Xaa₁₉ is: Tyr, Trp, or Phe;

Xaa₂₀ is: Leu, Phe, Tyr, or Trp;

10 Xaa₂₂ is: Gly, Glu, Asp, or Lys;

Xaa₂₅ is: Ala, Val, Ile, or Leu;

Xaa₂₇ is: Glu, Ile, or Ala;

Xaa₃₀ is: Ala or Glu

Xaa₃₃ is: Val, or Ile; and

15 Xaa₃₇ is: Gly, His, NH₂, or is absent.

70. The heterologous fusion protein of Claim 69 wherein the second polypeptide is human albumin.

20 71. The heterologous fusion protein of Claim 70 wherein the second polypeptide has the sequence of SEQ ID NO: 34.

72. The heterologous fusion protein of Claim 69 wherein the second polypeptide is an N-terminal fragment of albumin.

25 73. The heterologous fusion protein of any one of Claims 27, 28, or 29 wherein the GLP-1 compound comprises the sequence of formula IX [SEQ ID NO: 12]

30 Xaa₇-Xaa₈-Glu-Gly-Thr-Xaa₁₂-Thr-Ser-Asp-Xaa₁₆-Ser-
Xaa₁₈-Xaa₁₉-Xaa₂₀-Glu-Xaa₂₂-Gln-Ala-Xaa₂₅-Lys-Xaa₂₇-
Phe-Ile-Xaa₃₀-Trp-Leu-Xaa₃₃-Lys-Gly-Arg-Xaa₃₇
Formula IX (SEQ ID NO: 12)

wherein:

Xaa₇ is: L-histidine, D-histidine, desamino-histidine, 2-amino-histidine, β -hydroxy-histidine,

5 homohistidine, α -fluoromethyl-histidine, or α -methyl-histidine;

Xaa₈ is: Ala, Gly, Val, Leu, Ile, Ser, or Thr;

Xaa₁₂ is: Phe, Trp, or Tyr;

Xaa₁₆ is: Val, Trp, Ile, Leu, Phe, or Tyr;

10 Xaa₁₈ is: Ser, Trp, Tyr, Phe, Lys, Ile, Leu, Val;

Xaa₁₉ is: Tyr, Trp, or Phe;

Xaa₂₀ is: Leu, Phe, Tyr, or Trp;

Xaa₂₂ is: Gly, Glu, Asp, or Lys;

Xaa₂₅ is: Ala, Val, Ile, or Leu;

15 Xaa₂₇ is: Glu, Ile, or Ala;

Xaa₃₀ is: Ala or Glu

Xaa₃₃ is: Val, or Ile; and

Xaa₃₇ is: Gly, His, NH₂, or is absent.

20 74. The heterologous fusion protein of Claim 73 wherein the second polypeptide is the Fc portion of an Ig selected from the group consisting of: IgG1, IgG2, IgG3, IgG4, IgE, IgA, IgD, or IgM.

25 75. The heterologous fusion protein of Claim 73 or 74 wherein the second polypeptide is the Fc portion of an Ig selected from the group consisting of: IgG1, IgG2, IgG3, and IgG4.

30 76. The heterologous fusion protein of Claim 75 wherein the second polypeptide is the Fc portion of an IgG1 immunoglobulin.

- 119 -

77. The heterologous fusion protein of Claim 75 wherein the second polypeptide is the Fc portion of an IgG4 immunoglobulin.

5 78. The heterologous fusion protein of Claims 73 through 78 wherein the IgG is human.

10 79. The heterologous fusion protein of Claim 75 wherein the second polypeptide has the sequence of SEQ ID NO: 32.

15 80. The heterologous fusion protein of any one of Claims 69 through 79 wherein the GLP-1 compound is selected from the group consisting of Gly⁸-GLP-1(7-37), Val⁸-Tyr¹²-GLP-1(7-37), Val⁸-Tyr¹²-GLP-1(7-36), Val⁸-Trp¹²-GLP-1(7-37), Val⁸-Leu¹⁶-GLP-1(7-37), Val⁸-Val¹⁶-GLP-1(7-37), Val⁸-Tyr¹⁶-GLP-1(7-37), Gly⁸-Glu²²-GLP-1(7-37), Val⁸-Glu²²-GLP-1(7-37), Val⁸-Leu²⁵-GLP-1(7-37), Val⁸-Tyr¹²-Tyr¹⁶-GLP-1(7-37), Val⁸-Trp¹²-Glu²²-GLP-1(7-37), Val⁸-Tyr¹²-Glu²²-GLP-1(7-37), Val⁸-Tyr¹⁶-Phe¹⁹-GLP-1(7-37), Val⁸-Tyr¹⁶-Glu²²-GLP-1(7-37), Val⁸-Trp¹⁶-Glu²²-GLP-1(7-37), Val⁸-Leu¹⁶-Glu²²-GLP-1(7-37), Val⁸-Ile¹⁶-Glu²²-GLP-1(7-37), Val⁸-Phe¹⁶-Glu²²-GLP-1(7-37), Val⁸-Trp¹⁸-Glu²²-GLP-1(7-37), Val⁸-Tyr¹⁸-Glu²²-GLP-1(7-37), Val⁸-Phe¹⁸-Glu²²-GLP-1(7-37), Val⁸-Ile¹⁸-Glu²²-GLP-1(7-37), Val⁸-Lys¹⁸-Glu²²-GLP-1(7-37), Val⁸-Trp¹⁹-Glu²²-GLP-1(7-37), Val⁸-Phe¹⁹-Glu²²-GLP-1(7-37), Val⁸-Phe²⁰-Glu²²-GLP-1(7-37), Val⁸-Glu²²-Leu²⁵-GLP-1(7-37), Val⁸-Glu²²-Ile²⁵-GLP-1(7-37), Val⁸-Glu²²-Val²⁵-GLP-1(7-37), Val⁸-Glu²²-Ile²⁷-GLP-1(7-37), Val⁸-Glu²²-Ala²⁷-GLP-1(7-37), Val⁸-Glu²²-Ile³³-GLP-1(7-37), Val⁸-Asp⁹-Ile¹¹-Tyr¹⁶-Glu²²-GLP-1(7-37), Val⁸-Tyr¹⁶-Trp¹⁹-Glu²²-GLP-1(7-37), Val⁸-Trp¹⁶-Glu²²-Val²⁵-Ile³³-GLP-1(7-37), Val⁸-Trp¹⁶-Glu²²-Ile³³-GLP-1(7-37), Val⁸-Glu²²-Val²⁵-Ile³³-GLP-1(7-37), Val⁸-Trp¹⁶-

- 120 -

Glu²²-Val²⁵- GLP-1 (7-37) , and Val⁸-Cys¹⁶-Lys²⁶-GLP-1 (7-37) .

1/17

Fig. 1

2/17

Fig. 2

5	10	15
Asp	Ala	His
Lys	Ser	Glu
20	25	30
Phe	Lys	Ala
Leu	Val	Leu
Ile	Ala	Phe
40	45	50
Glu	Asp	His
His	Val	Lys
Leu	Val	Asn
55	60	65
Ala	Asp	Glu
Glu	Ser	Ala
Asn	Cys	Asp
75	80	85
Lys	Leu	Cys
Thr	Val	Ala
Ala	Thr	Leu
95	100	105
Cys	Ala	Lys
Gln	Glu	Pro
110	115	120
Asn	Pro	Asn
Leu	Pro	Arg
130	135	140
Phe	His	Asp
Asn	Glu	Glu
145	150	155
Arg	His	Pro
Tyr	Phe	Tyr
165	170	175
Ala	Ala	Phe
Thr	Glu	Cys
185	190	195
Lys	Leu	Asp
Glu	Leu	Arg
200	205	210
Lys	Cys	Ala
Ala	Ser	Leu
220	225	230
Ala	Arg	Leu
Ser	Gln	Arg
235	240	245
Val	Thr	Asp
Leu	Thr	Lys
255	260	265
Cys	Ala	Asp
Asp	Arg	Ala
275	280	285
Ile	Ser	Ser
Lys	Leu	Lys
290	295	300
Cys	Ile	Ala
Glu	Val	Glu
310	315	320
Ala	Asp	Phe
325	330	335
Val	Phe	Leu
Gly	Met	Phe
345	350	355
Val	Val	Leu
Leu	Leu	Arg
365	370	375
Cys	Ala	Ala
Ala	Asp	Pro
380	385	390
Pro	Leu	Val
Glu	Glu	Pro
400	405	410
Asn	Leu	Gly
Glu	Glu	Tyr
415	420	425
Val	Pro	Gln
Gln	Val	Ser
435	440	445
Val	Gly	Ser
Lys	Cys	Cys
455	460	465
Asp	Tyr	Leu
Tyr	Leu	Ser
470	475	480
Val	Ser	Asp
Arg	Val	Thr
Lys	Cys	Cys
485	490	495
Val	Ser	Asp
Arg	Val	Asn
Arg	Arg	Pro

3/17

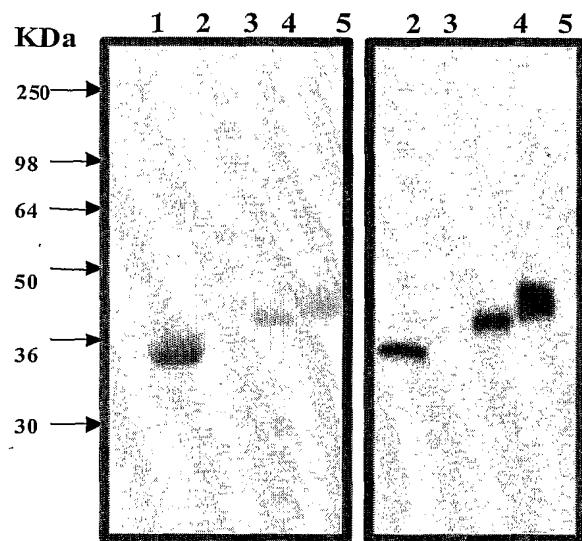
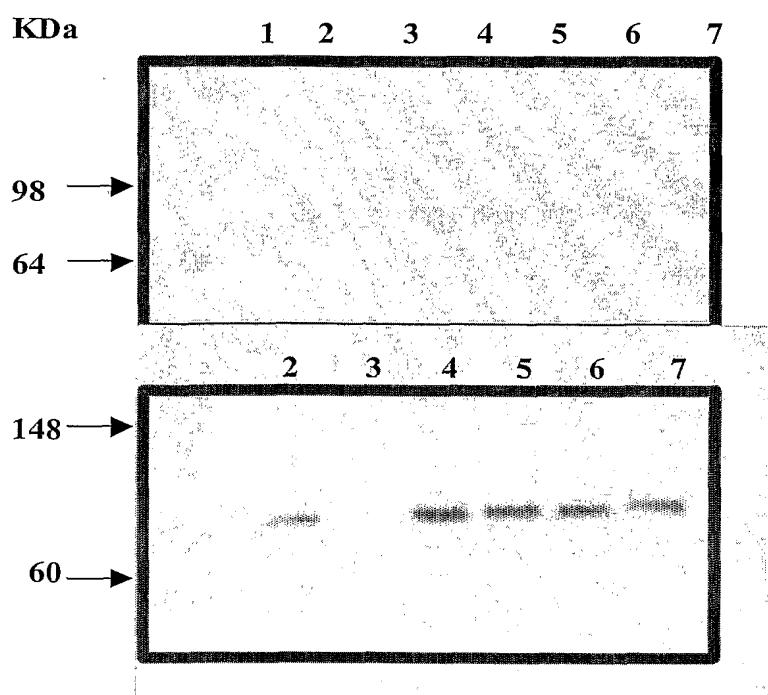
Fig. 2 Continued

490	495	500
Cys Phe Ser Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys Glu Phe Asn Ala		
505	510	515
Glu Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu Arg Gln		
525	530	535
Ile Lys Lys Gln Thr Ala Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr		
545	550	555
Lys Glu Gln Leu Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys		
560	565	570
Cys Lys Ala Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val		
580	585	
Ala Ala Ser Gln Ala Ala Leu Gly Leu		

[SEQ ID NO: 34]

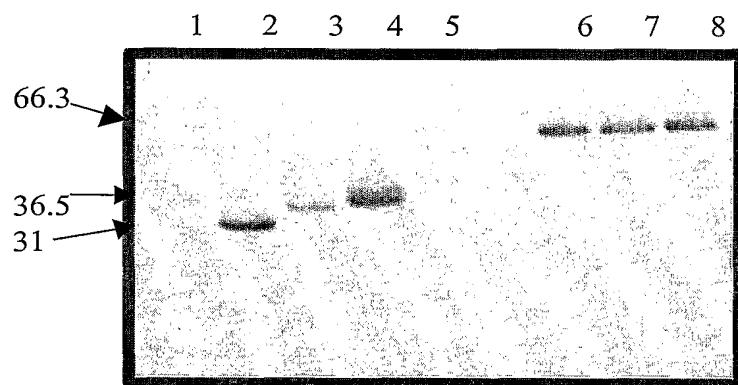
4/17

Fig. 3

A.**B.**

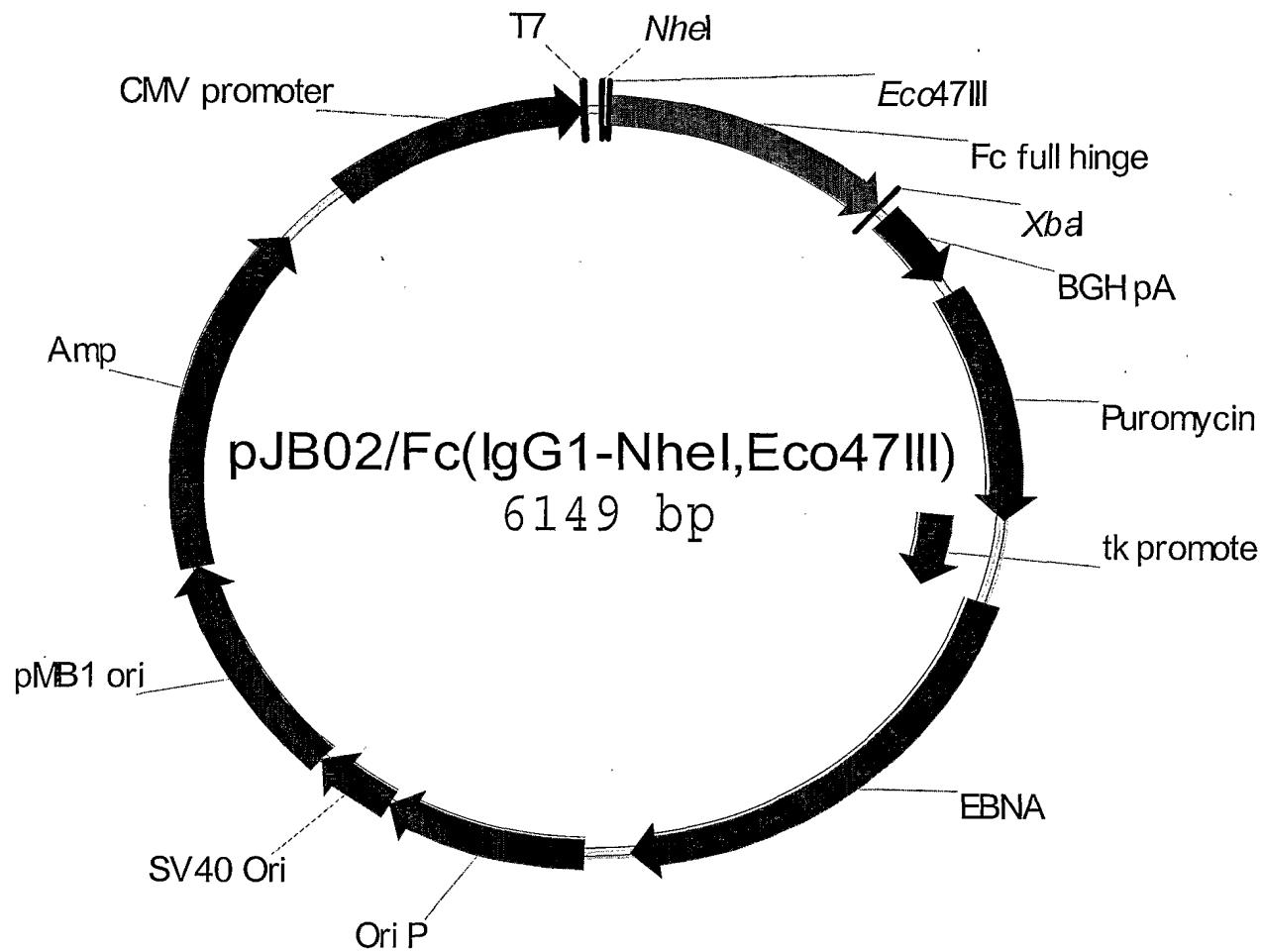
5/17

Fig. 4



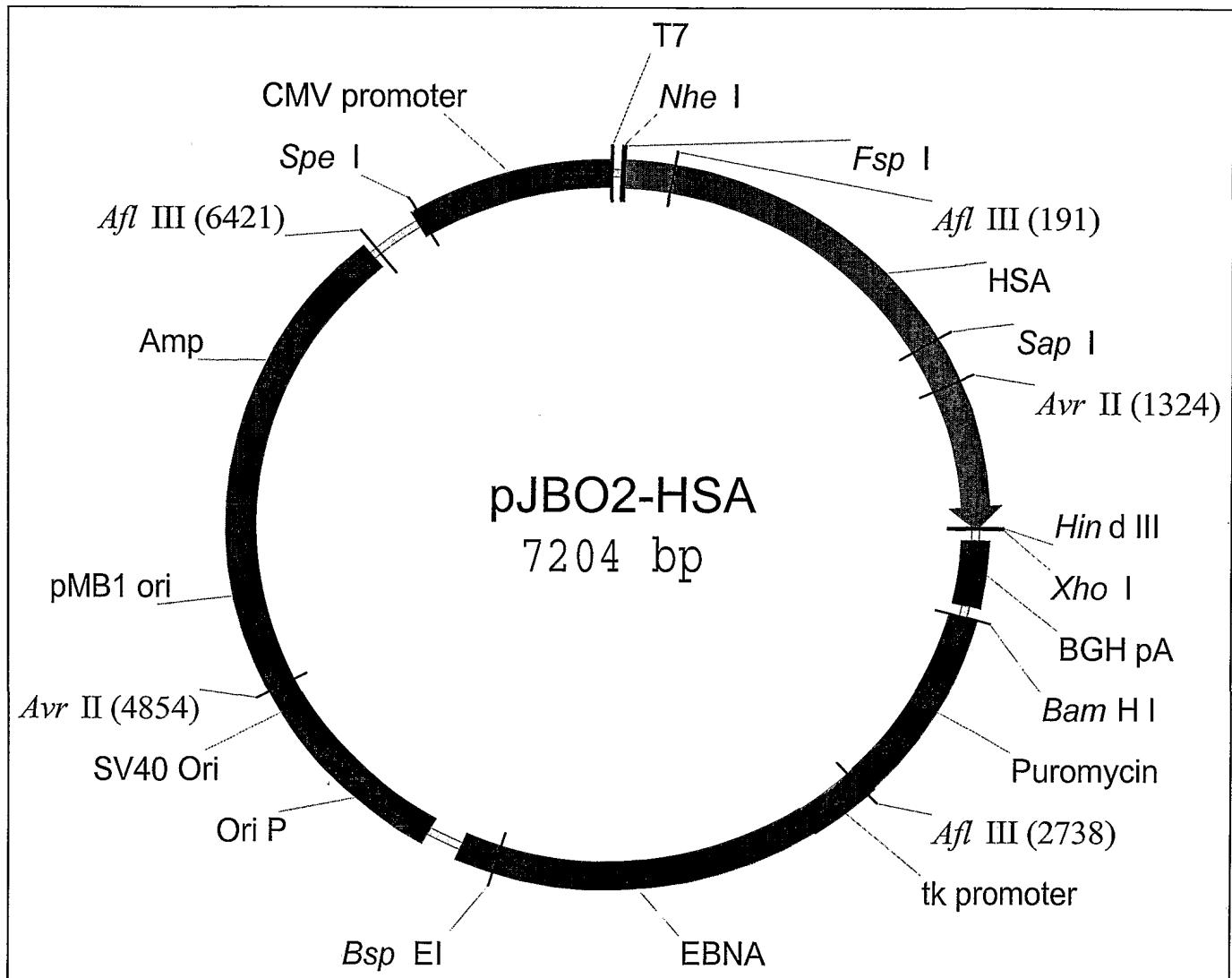
6/17

Fig. 5



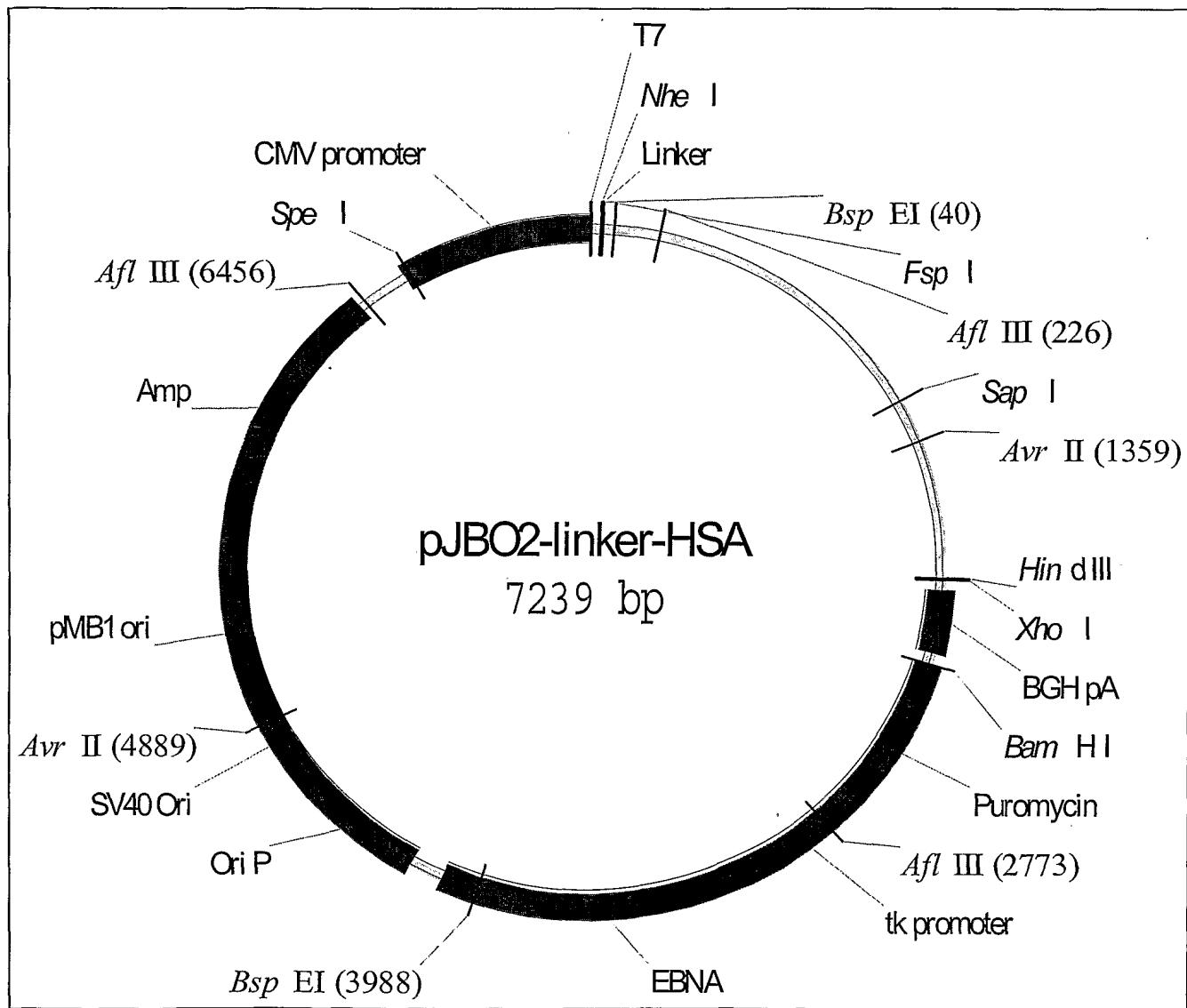
7/17

Fig. 6



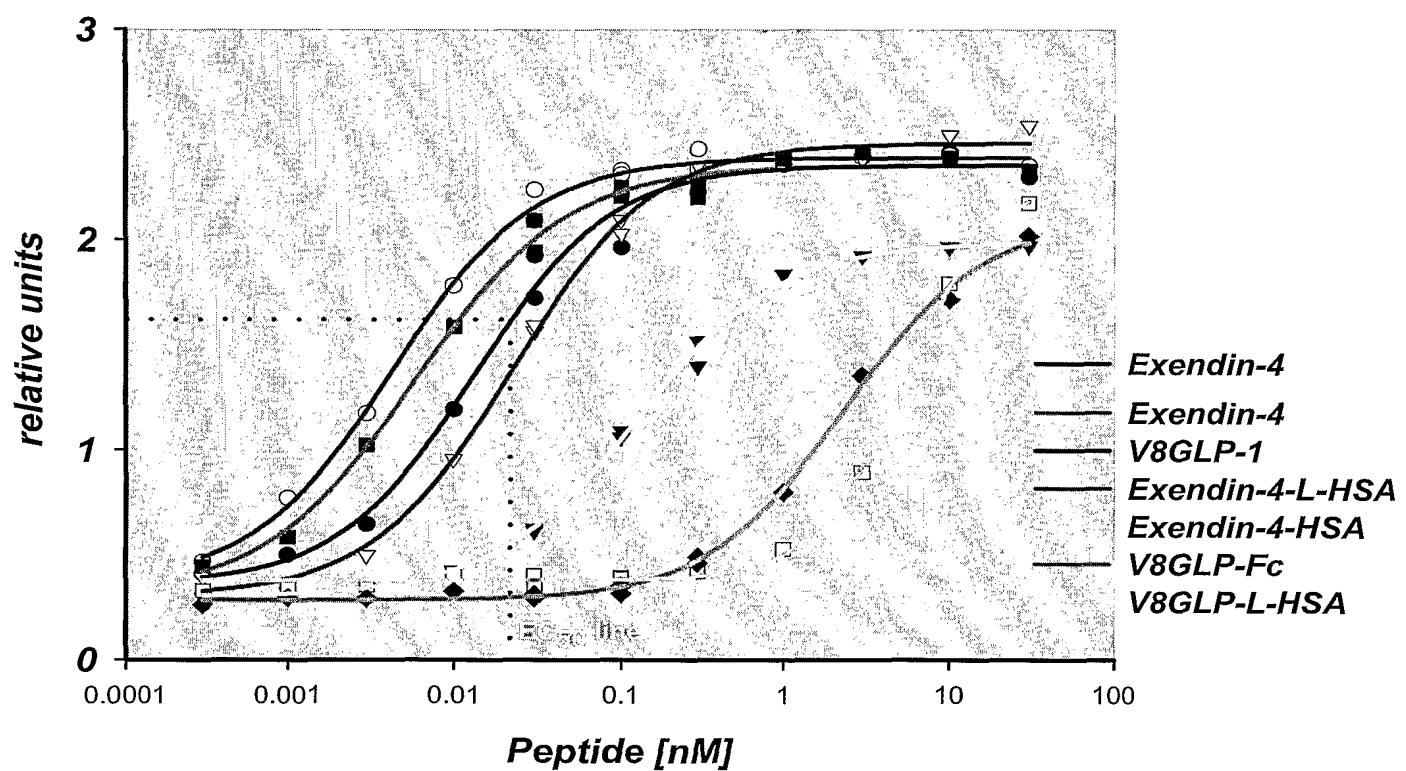
8/17

Fig. 7



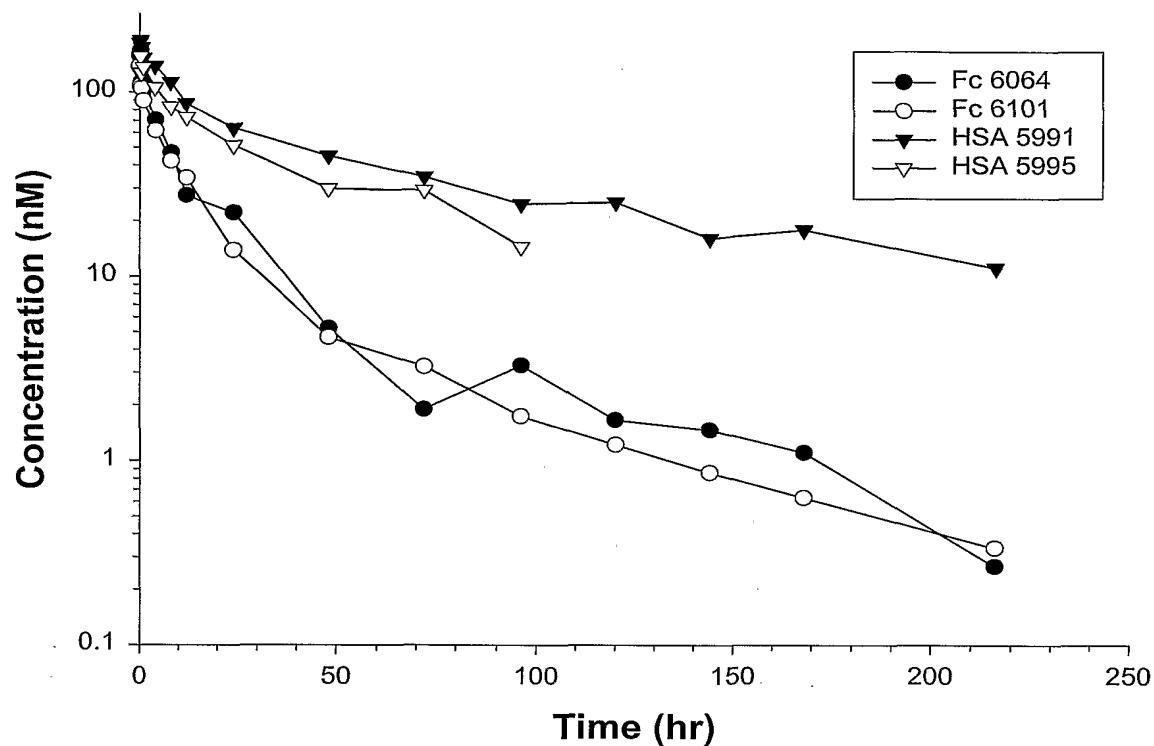
9/17

Fig. 8



10/17

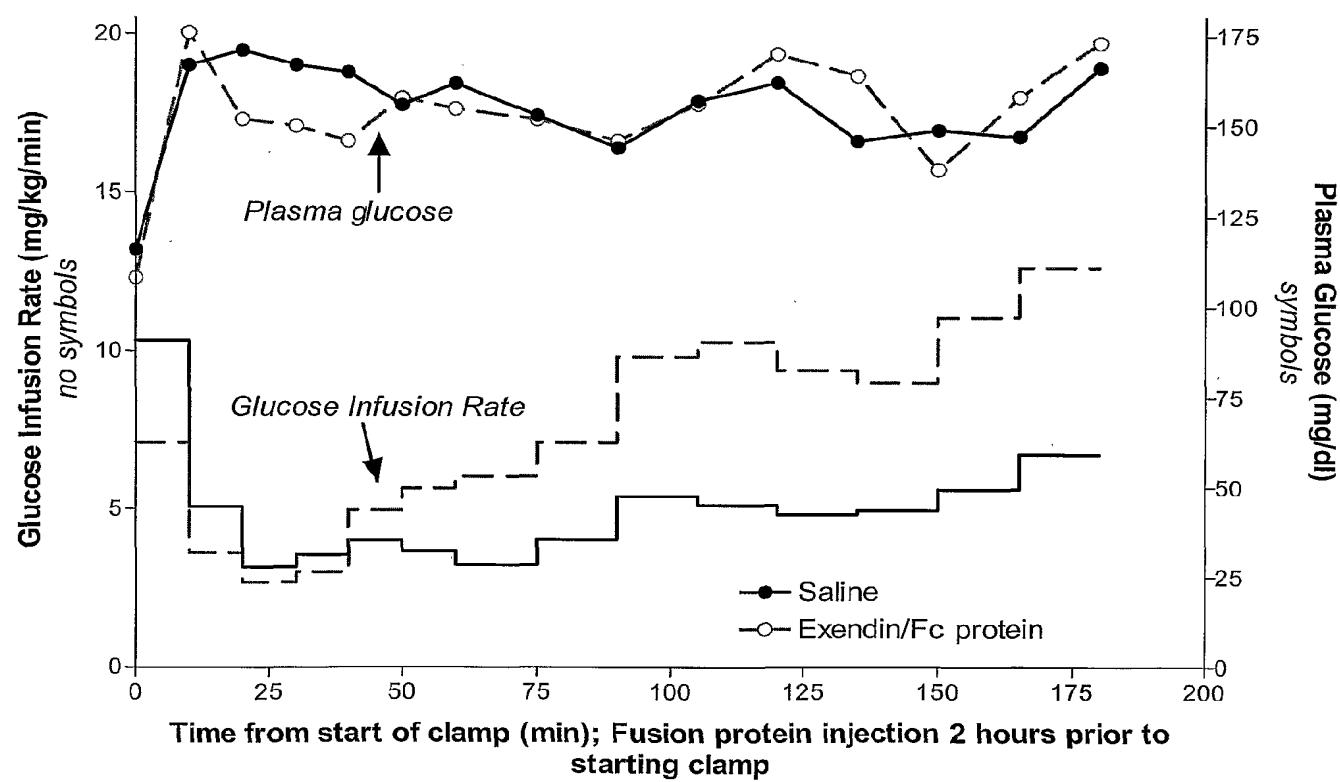
Fig. 9



11/17

Fig. 10

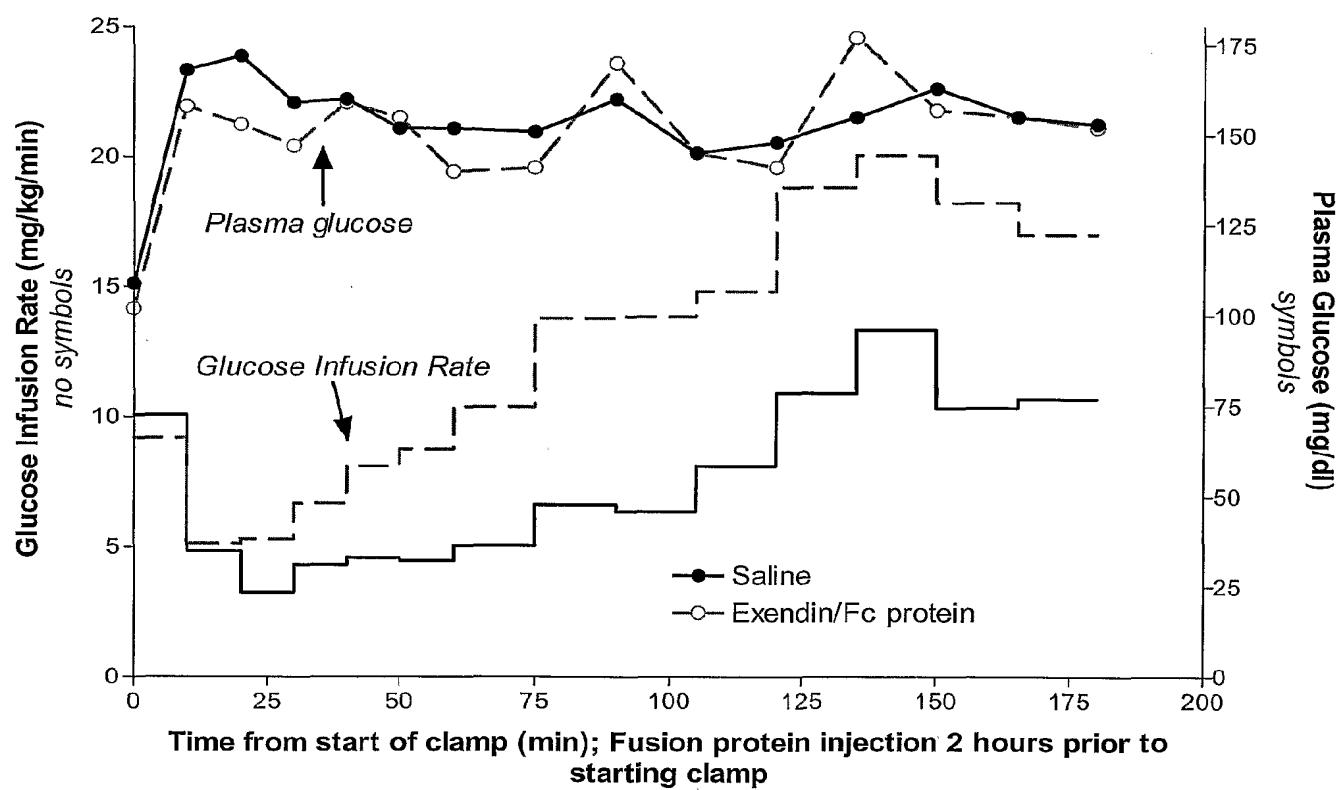
A.



12/17

Fig. 10

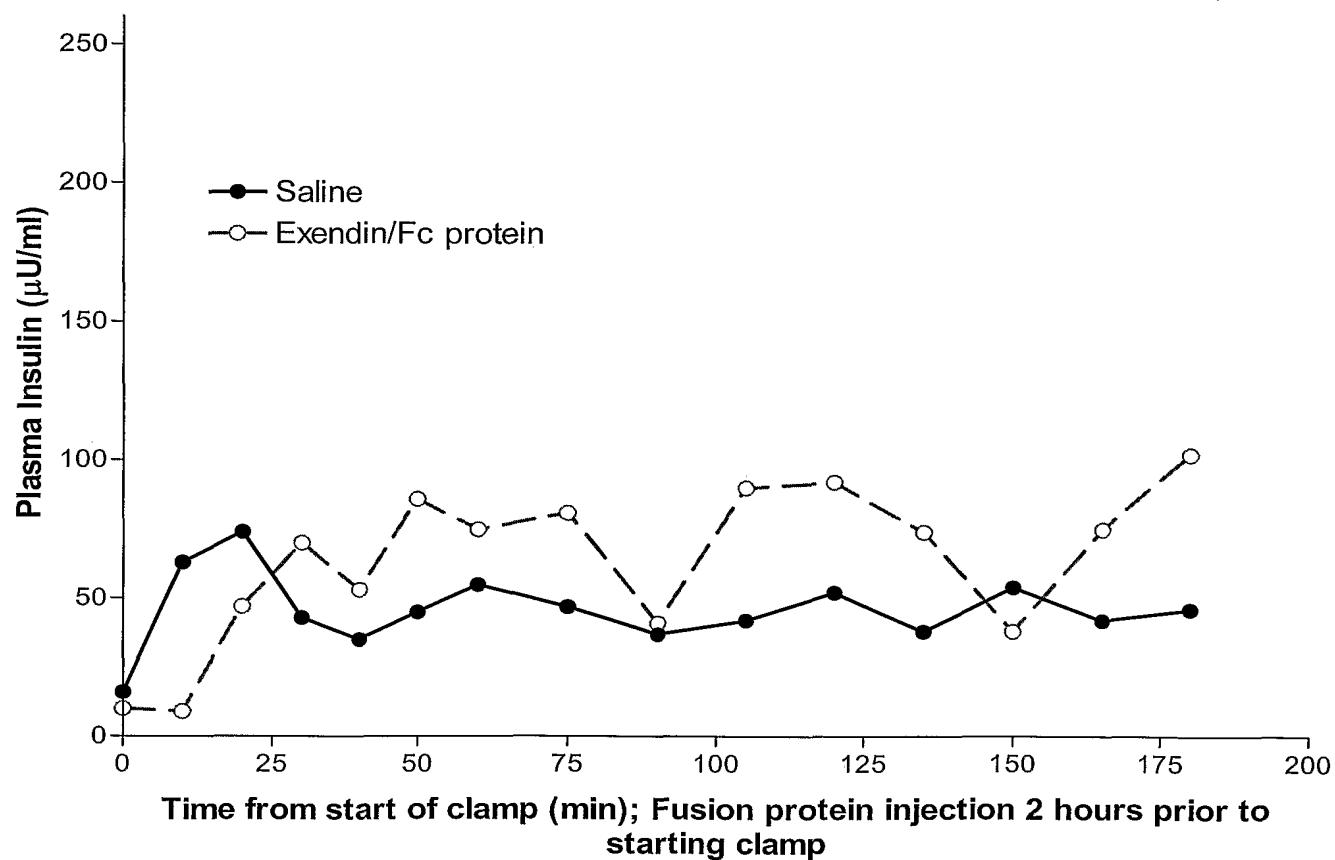
B.



13/17

Fig. 11

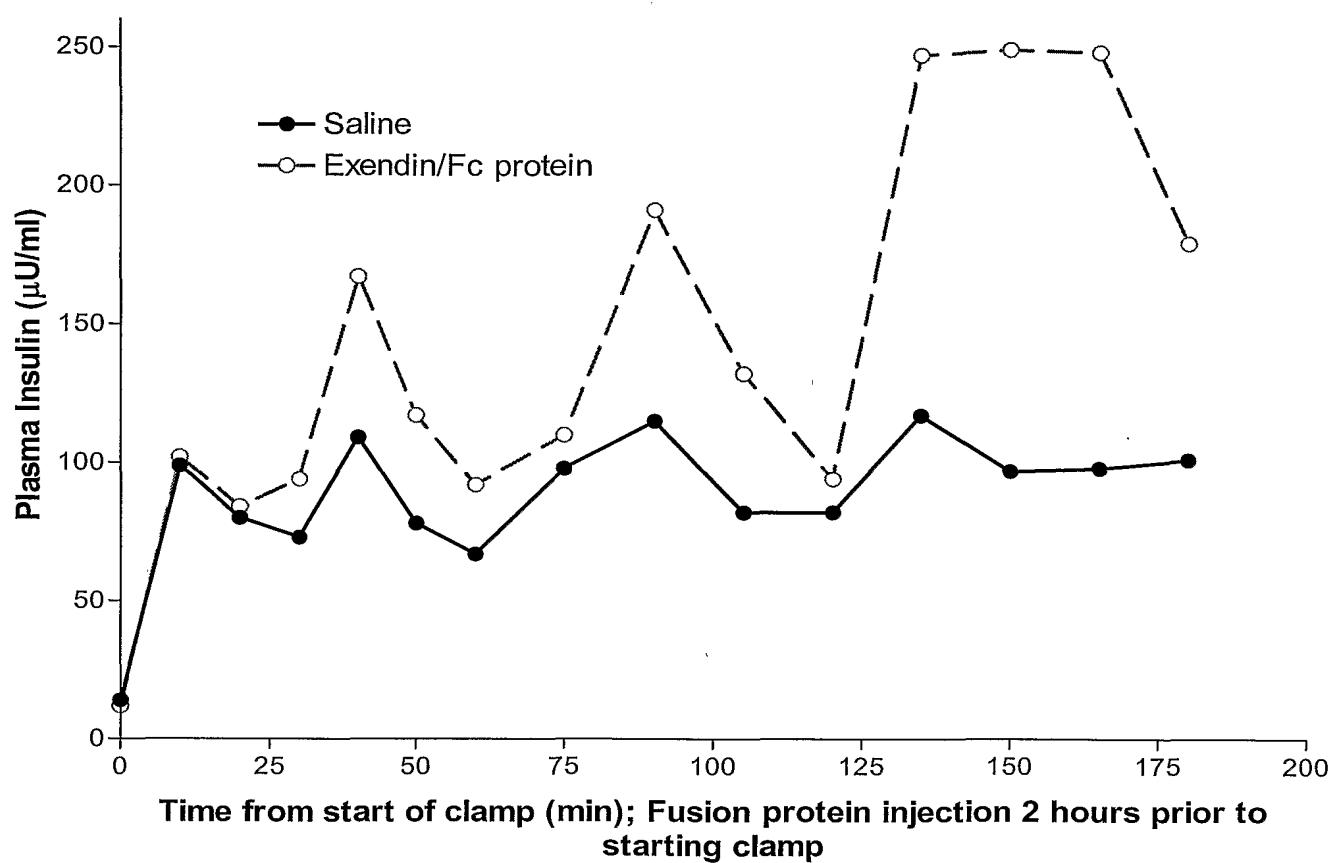
A.



14/17

Fig. 11

B.



15/17

Fig. 12

1
GAGCCCAAATCTTGTGACAAAACACACATGCCACCGTGCCCAGCACC
50
TGAACCTCCTGGGGGACCGTCAGTCTTCCCTCTTCCCCAAAACCCAAGG
100
ACACCCTCATGATCTCCGGACCCCTGAGGTACATGCGTGGTGGAC
150
GTGAGCCACGAAGACCCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGT
200
GGAGGTGCATAATGCCAAGACAAAGCCGGAGGAGCAGTACAACAGCA
250
CGTACCGTGTGGTCAGCGTCCTCACCGTCTGCACCAGGACTGGCTGAAT
300
GGCAAGGAGTACAAGTCAAGGTCTCCAACAAAGCCCTCCCAGCCCCAT
350
CGAGAAAACCATCTCAAAGCCAAGGGCAGCCCCGAGAACCAACAGGTGT
400
ACACCCTGCCCATCCGGGAGGAGATGACCAAGAACCAAGGTCAAGCCTG
450
ACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATGCCGTGGAGTGGGA
500
GAGCAATGGGCAGCCGGAGAACAAACTACAAGACCACGCCTCCGTGCTGG
550
ACTCCGACGGCTCCTTCTTCCTCTATAGCAAGCTCACCGTGGACAAGAGC
600
AGGTGGCAGCAGGGAACGTCTCTCATGCTCCGTGATGCATGAGGCTCT
650
GCACAACCAACTACACGCAGAAGAGCCTCTCCGTCTCCGGTAAATGAT
700
AGT [SEQ ID NO: 33]

16/17

Fig. 13

1
GATGCGCACAAAGAGTGAGGTTGCTCATCGTTAAAGATTGGGAGAAGA
50
AAATTCAAAGCCTGGTGTGATTGCCTTGCTCAGTATCTCAGCAGT
100
GTCCATTTGAAGATCATGTAAAATTAGTGAATGAAGTAAC TGAAATTGCA
150
AAAACATGTGTTGCTGATGAGTCAGCTGAAAATTGTGACAAATCACTTCA
200
TACCCCTTTGGAGACAAATTATGCACAGTTGCAACTCTCGTGAACACCT
250
ATGGTGAATGGCTGACTGCTGTGCAAAACAAGAACCTGAGAGAAATGAA
300
TGCTTCTTGCAACACAAAGATGACAACCCAAACCTCCCCGATTGGTGAG
350
ACCAGAGGTTGATGTGATGTGCACTGCTTTCATGACAATGAAGAGACAT
400
TTTGAAAAAAACTTATATGAAATTGCCAGAAGACATCCTTACTTTAT
450
GCCCGGAACTCCTTCTTGCTAAAGGTATAAGCTGCTTTACAGA
500
ATGTTGCCAAGCTGCTGATAAGCTGCCTGCCTGTTGCCAAAGCTCGATG
550
AACTCGGGATGAAGGGAAGGCTTCGTCGCCAAACAGAGACTCAAGTGT
600
GCCAGTCTCCAAAATTGGAGAAAGAGCTTCAAAGCATGGCAGTAGC
650
TCGCCTGAGCCAGAGATTCCAAAGCTGAGTTGCAGAAGTTCCAAGT
700
TAGTGACAGATCTTACCAAAGTCCACACGGAATGCTGCCATGGAGATCTG
750
CTTGAATGTGCTGATGACAGGGCGGACCTGCCAAGTATATCTGTGAAAA
800
TCAAGATTGATCTCCAGTAAACTGAAGGAATGCTGTGAAAAACCTCTGT
850
TGGAAAAATCCCACTGCATTGCCAGTGGAAAATGATGAGATGCCCTGCT
900
GACTTGCCTTCATTAGCTGCTGATTTGTTGAAAGTAAGGATGTTGCAA
950
AAACTATGCTGAGGCAAAGGATGTCTCCTGGCATGTTTGATGAAT
1000
ATGCAAGAAGGCATCCTGATTACTCTGCTGCTGCTGAGACTTGCC

17/17

Fig. 13 Continued

1050 AAGACATATGAAACCACCTAGAGAAGTGCTGTGCCGCTGCAGATCCTCA
1100 TGAATGCTATGCCAAAGTGTGATGAATTAAACCTTTGTGGAAGAGC
1150 CTCAGAATTAAATCAAACAAAATTGTGAGCTTTGAGCAGCTGGAGAG
1200 TACAAATTCCAGAATGCGCTATTAGTTGTTACACCAAGAAAGTACCCCA
1250 AGTGTCAACTCCAACCTTTGTAGAGGTCTCAAGAACCTAGGAAAAGTGG
1300 GCAGCAAATGTTGTAACATCCTGAAGCAAAAGAATGCCCTGTGCAGAA
1350 GACTATCTATCCGTGGCCTGAACCAGTTATGTGTGTTGCATGAGAAAAC
1400 GCCAGTAAGTGACAGAGTCACCAAATGCTGCACAGAACATCCTGGTGAACA
1450 GGCGACCATGCTTTCAGCTCTGGAAGTCGATGAAACATACGTTCCAAA
1500 GAGTTAACATGCTGAAACATTCACCTTCATGCAGATATGCACACTTC
1550 TGAGAAGGAGAGACAAATCAAGAAACAAACTGCACTTGTGAGCTCGTGA
1600 AACACAAGCCCAAGGCAACAAAAGAGCAACTGAAAGCTGTTATGGATGAT
1650 TTCGAGCTTTGTAGAGAAGTGCTGCAAGGCTGACGATAAGGAGACCTG
1700 CTTGCCGAGGAGGGTAAAAACTTGTGCTGCAAGTCAAGCTGCCTTAG
1750 GCTTATAATGAC [SEQ ID NO: 35]

X-13991.ST25.txt
SEQUENCE LISTING

<110> Eli Lilly and Company

<120> GLP-1 FUSION PROTEINS

<130> X-13991

<150> US 60/251,954

<151> 2000-06-12

<160> 35

<170> PatentIn version 3.1

<210> 1

<211> 31

<212> PRT

<213> Homo sapiens

<400> 1

His Ala Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly
20 25 30

<210> 2

<211> 39

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<220>

<221> MISC_FEATURE

x-13991.ST25.txt

<222> (2)..(2)

<223> Xaa at position 2 is Ala, Gly, Ser, Thr, Leu, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (3)..(3)

<223> Xaa at position 3 is Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (5)..(5)

<223> Xaa at position 5 is Thr, Ala, Gly, Ser, Leu, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (8)..(8)

<223> Xaa at position 8 is Ser, Ala, Gly, Thr, Leu, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (10)..(10)

<223> Xaa at position 10 is Val, Ala, Gly, Ser, Thr, Leu, Ile, Tyr, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (11)..(11)

<223> Xaa at position 11 is Ser, Ala, Gly, Thr, Leu, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (12)..(12)

x-13991.ST25.txt
<223> Xaa at position 12 is Ser, Ala, Gly, Thr, Leu, Ile, Val, Glu, Asp, Lys, Trp or Tyr;

<220>

<221> MISC_FEATURE

<222> (13)..(13)

<223> Xaa at position 13 is Tyr, Phe, Trp, Glu, Asp, Gln, or Lys;

<220>

<221> MISC_FEATURE

<222> (14)..(14)

<223> Xaa at position 14 is Leu, Ala, Gly, Ser, Thr, Ile, Val, Glu, Asp, Met, Lys, Trp or Tyr;

<220>

<221> MISC_FEATURE

<222> (15)..(15)

<223> Xaa at position 15 is Glu, Asp, or Lys;

<220>

*

<221> MISC_FEATURE

<222> (16)..(16)

<223> Xaa at position 16 is Gly, Ala, Ser, Thr, Leu, Ile, Val, Glu, Asp, Trp or Lys;

<220>

<221> MISC_FEATURE

<222> (17)..(17)

<223> Xaa at position 17 is Gln, Asn, Arg, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (18)..(18)

<223> Xaa at position 18 is Ala, Gly, Ser, Thr, Leu, Ile, Val, Arg, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (19)..(19)

<223> Xaa at position 19 is Ala, Gly, Ser, Thr, Leu, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (20)..(20)

<223> Xaa at position 20 is Lys, Arg, Gln, Glu, Asp, or His;

<220>

<221> MISC_FEATURE

<222> (21)..(21)

<223> Xaa at position 21 is Leu, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (24)..(24)

<223> Xaa at position 24 is Ala, Gly, Ser, Thr, Leu, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (25)..(25)

<223> Xaa at position 25 is Trp, Phe, Tyr, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (26)..(26)

X-13991.ST25.txt

<223> Xaa at position 26 is Leu, Gly, Ala, Ser, Thr, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (27)..(27)

<223> Xaa at position 27 is Val, Gly, Ala, Ser, Thr, Leu, Ile, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (28)..(28)

<223> Xaa at position 28 is Asn, Lys, Arg, Glu, Asp, or His;

<220>

<221> MISC_FEATURE

<222> (29)..(29)

<223> Xaa at position 29 is Gly, Ala, Ser, Thr, Leu, Ile, Val, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

<222> (30)..(30)

<223> Xaa at position 30 is Gly, Arg, Lys, Glu, Asp, or His;

<220>

<221> MISC_FEATURE

<222> (31)..(31)

<223> Xaa at position 31 is Pro, Gly, Ala, Ser, Thr, Leu, Ile, Val, Glu, Asp, or Lys, or is deleted;

<220>

<221> MISC_FEATURE

x-13991.ST25.txt

<222> (32)..(32)

<223> Xaa at position 32 is Ser, Arg, Lys, Glu, Asp, or His, or is deleted;

<220>

<221> MISC_FEATURE

<222> (33)..(33)

<223> Xaa at position 33 is Ser, Arg, Lys, Glu, Asp, or His, or is deleted;

<220>

<221> MISC_FEATURE

<222> (34)..(34)

<223> Xaa at position 34 is Gly, Asp, Glu, or Lys, or is deleted;

<220>

<221> MISC_FEATURE

<222> (35)..(35)

<223> Xaa at position 35 is Ala, Phe, Trp, Tyr, Glu, Asp, or Lys, or is deleted;

<220>

<221> MISC_FEATURE

<222> (36)..(36)

<223> Xaa at position 36 is Ser, Pro, Lys, Glu, or Asp, or is deleted;

<220>

<221> MISC_FEATURE

<222> (37)..(37)

<223> Xaa at position 37 is Ser, Pro, Glu, Asp, or Lys, or is deleted;

<220>

<221> MISC_FEATURE

X-13991.ST25.txt

<222> (38)..(38)

<223> Xaa at position 38 is Gly, Pro, Glu, Asp, or Lys, or is deleted;

<220>

<221> MISC_FEATURE

<222> (39)..(39)

<223> Xaa at position 39 is Ala, Ser, Val, Glu, Asp, or Lys, or is deleted;

<400> 2

His Xaa Xaa Gly Xaa Phe Thr Xaa Asp Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15Xaa Xaa Xaa Xaa Xaa Phe Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
35

<210> 3

<211> 32

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa at position 1 is L-histidine, D-histidine, or is deleted.

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> Xaa at position 2 is Gly, Ala, Val, Leu, Ile, Ser, or Thr;

<220>

X-13991.ST25.txt

<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa at position 3 is Thr, Ser, Arg, Lys, Trp, Phe, Tyr, Glu, or His;

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa at position 5 is Asp, Glu, Arg, Thr, Ala, Lys, or His;

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> Xaa at position 6 is His, Trp, Phe, or Tyr;

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> Xaa at position 10 is Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Tyr, Glu, or Ala;

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> Xaa at position 12 is His, Pro, Asp, Glu, Arg, Ser, Ala, or Lys;

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> Xaa at position 13 is Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

<220>

X-13991.ST25.txt

<221> MISC_FEATURE
<222> (17)..(17)
<223> Xaa at position 17 is His, Asp, Lys, Glu, Gln, or Arg;

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> Xaa at position 18 is Glu, Arg, Ala, or Lys;

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> Xaa at position 20 is Trp, Tyr, Phe, Asp, Lys, Glu, or His;

<220>
<221> MISC_FEATURE
<222> (21)..(21)
<223> Xaa at position 21 is Ala, Glu, His, Phe, Tyr, Trp, Arg, or Lys;

<220>
<221> MISC_FEATURE
<222> (24)..(24)
<223> Xaa at position 24 is Ala, Glu, Asp, Ser, or His;

<220>
<221> MISC_FEATURE
<222> (25)..(25)
<223> Xaa at position 25 is Asp, Glu, Ser, Thr, Arg, Trp, or Lys;

<220>
<221> MISC_FEATURE

X-13991.ST25.txt

<222> (27)..(27)

<223> Xaa at position 27 is Asp, Arg, Val, Lys, Ala, Gly, or Glu;

<220>

<221> MISC_FEATURE

<222> (28)..(28)

<223> Xaa at position 28 is Glu, Lys, or Asp;

<220>

<221> MISC_FEATURE

<222> (29)..(29)

<223> Xaa at position 29 is Thr, Ser, Lys, Arg, Trp, Tyr, Phe, Asp, Gly, Pro, His, or Glu;

<220>

<221> MISC_FEATURE

<222> (30)..(30)

<223> Xaa at position 30 is Thr, Ser, Asp, Trp, Tyr, Phe, Arg, Glu, or His;

<220>

<221> MISC_FEATURE

<222> (31)..(31)

<223> Xaa at position 31 is Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, or is deleted.

<220>

<221> MISC_FEATURE

<222> (32)..(32)

<223> Xaa at position 31 is Pro or is deleted.

<400> 3

Xaa Xaa Xaa Gly Xaa Xaa Thr Ser Asp Xaa Ser Xaa Xaa Leu Glu Gly
1 5 10 15

Xaa Xaa Ala Xaa Xaa Xaa Phe Ile Xaa Xaa Leu Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

X-13991.ST25.txt

<210> 4
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic construct
<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> Xaa at position 1 is L-histidine, D-histidine, or is deleted.

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> Xaa at position 2 is Gly, Ala, Val, Leu, Ile, Ser, or Thr;

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa at position 5 is Asp, Glu, Arg, Thr, Ala, Lys, or His;

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> Xaa at position 6 is His, Trp, Phe, or Tyr;

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> Xaa at position 10 is Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Glu, or Ala;

x-13991.ST25.txt

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> Xaa at position 16 is Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> Xaa at position 17 is His, Asp, Lys, Glu, or Gln;

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> Xaa at position 18 is Glu, His, Ala, or Lys;

<220>
<221> MISC_FEATURE
<222> (19)..(19)
<223> Xaa at position 19 is Asp, Lys, Glu, or His;

<220>
<221> MISC_FEATURE
<222> (21)..(21)
<223> Xaa at position 21 is Ala, Glu, His, Phe, Tyr, Trp, Arg, or Lys;

<220>
<221> MISC_FEATURE
<222> (24)..(24)
<223> Xaa at position 24 is Ala, Glu, Asp, Ser, or His;

x-13991.ST25.txt

<220>

<221> MISC_FEATURE

<222> (27)..(27)

<223> Xaa at position 27 is Asp, Arg, Val, Lys, Ala, Gly, or Glu;

<220>

<221> MISC_FEATURE

<222> (28)..(28)

<223> Xaa at position 28 is Glu, Lys, or Asp;

<220>

<221> MISC_FEATURE

<222> (29)..(29)

<223> Xaa at position 29 is Thr, Ser, Lys, Arg, Trp, Tyr, Phe, Asp, Gly, Pro, His, or Glu;

<220>

<221> MISC_FEATURE

<222> (30)..(30)

<223> Xaa at position 30 is Arg, Glu, or His;

<220>

<221> MISC_FEATURE

<222> (31)..(31)

<223> Xaa at position 31 is Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, or is deleted.

<220>

<221> MISC_FEATURE

<222> (32)..(32)

<223> Xaa at position 32 is Pro, or is deleted.

<400> 4

Xaa Xaa Glu Gly Xaa Xaa Thr Ser Asp Xaa Ser Ser Tyr Leu Glu Xaa
1 5 10 15

X-13991.ST25.txt

Xaa Xaa Xaa Lys Xaa Phe Ile Xaa Trp Leu Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

<210> 5

<211> 32

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa at position 1 is L-histidine, D-histidine, or is deleted.

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> Xaa at position 2 is Gly, Ala, Val, Leu, Ile, Ser, Met, or Thr;

<220>

<221> MISC_FEATURE

<222> (6)..(6)

<223> Xaa at position 6 is His, Trp, Phe, or Tyr;

<220>

<221> MISC_FEATURE

<222> (10)..(10)

<223> Xaa at position 10 is Leu, Ser, Thr, Trp, His, Phe, Asp, Val, Glu, or Ala;

<220>

<221> MISC_FEATURE

<222> (16)..(16)

X-13991.ST25.txt

<223> Xaa at position 16 is Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

<220>

<221> MISC_FEATURE

<222> (17)..(17)

<223> Xaa at position 17 is His, Asp, Lys, Glu, or Gln;

<220>

<221> MISC_FEATURE

<222> (20)..(20)

<223> Xaa at position 20 is Asp, Lys, Glu, or His;

<220>

<221> MISC_FEATURE

<222> (24)..(24)

<223> Xaa at position 24 is Ala, Glu, Asp, Ser, or His;

<220>

<221> MISC_FEATURE

<222> (29)..(29)

<223> Xaa at position 29 is Thr, Ser, Lys, Arg, Trp, Tyr, Phe, Asp, Gly, Pro, His, or Glu;

<220>

<221> MISC_FEATURE

<222> (31)..(31)

<223> Xaa at position 31 is Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, or is deleted.

<220>

<221> MISC_FEATURE

<222> (32)..(32)

<223> Xaa at position 32 is Pro or is deleted.

X-13991.ST25.txt

<400> 5

Xaa Xaa Glu Gly Thr Xaa Thr Ser Asp Xaa Ser Ser Tyr Leu Glu Xaa
1 5 10 15

Xaa Ala Ala Xaa Glu Phe Ile Xaa Trp Leu Val Lys Xaa Arg Xaa Xaa
20 25 30

<210> 6

<211> 32

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa at position 1 is L-histidine, D-histidine, or is deleted.

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> Xaa at position 2 is Gly, Ala, Val, Leu, Ile, Ser, or Thr;

<220>

<221> MISC_FEATURE

<222> (16)..(16)

<223> Xaa at position 16 is Gly, Asp, Glu, Gln, Asn, Lys, Arg, or Cys;

<220>

<221> MISC_FEATURE

<222> (17)..(17)

<223> Xaa at position 17 is His, Asp, Lys, Glu, or Gln;

<220>

<221> MISC_FEATURE

X-13991.ST25.txt

<222> (18)..(18)

<223> Xaa at position 18 is Ala, Glu, His, Phe, Tyr, Trp, Arg, or Lys;

<220>

<221> MISC_FEATURE

<222> (24)..(24)

<223> Xaa at position 24 is Ala, Glu, Asp, Ser, or His;

<220>

<221> MISC_FEATURE

<222> (31)..(31)

<223> Xaa at position 31 is Lys, Arg, Thr, Ser, Glu, Asp, Trp, Tyr, Phe, His, Gly, Gly-Pro, or is deleted.

<220>

<221> MISC_FEATURE

<222> (32)..(32)

<223> Xaa at position 32 is Pro or is deleted.

<400> 6

Xaa Xaa Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Xaa
1 5 10 15

Xaa Xaa Ala Lys Glu Phe Ile Xaa Trp Leu Val Lys Gly Arg Xaa Xaa
20 25 30

<210> 7

<211> 31

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<220>

<221> MISC_FEATURE

X-13991.ST25.txt

<222> (1)..(1)

<223> Xaa at position 1 is L-histidine, D-histidine, or is deleted.

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> Xaa at position 2 is Ala, Gly, Val, Thr, Ile, and alpha-methyl-Ala;

<220>

<221> MISC_FEATURE

<222> (15)..(15)

<223> Xaa at position 15 is Glu, Gln, Ala, Thr, Ser, and Gly;

<220>

<221> MISC_FEATURE

<222> (21)..(21)

<223> Xaa at position 21 is Glu, Gln, Ala, Thr, Ser, and Gly;

<400> 7

Xaa Xaa Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Xaa Gly
1 5 10 15

Gln Ala Ala Lys Xaa Phe Ile Ala Trp Leu Val Lys Gly Arg Gly
20 25 30

<210> 8

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<220>

<221> MISC_FEATURE

<222> (19)..(19)

X-13991.ST25.txt

<223> Xaa at position 19 is Lys or Arg;

<220>

<221> MOD_RES

<222> (27)..(27)

<223> ACETYLATION

<220>

<221> MISC_FEATURE

<222> (30)..(30)

<223> Xaa at position 30 is Gly;

<220>

<221> MOD_RES

<222> (30)..(30)

<223> AMIDATION

<400> 8

Ala Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly Gln
1 5 10 15Ala Ala Xaa Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Xaa
20 25 30

<210> 9

<211> 39

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 9

His Ser Asp Gly Thr Phe Thr Ser Asp Leu Ser Lys Gln Met Glu Glu
1 5 10 15Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn Gly Gly Pro Ser
20 25 30

x-13991.ST25.txt

Ser Gly Ala Pro Pro Pro Ser
35

<210> 10
<211> 39
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic construct
<400> 10

His Gly Glu Gly Thr Phe Thr Ser Asp Leu Ser Lys Gln Met Glu Glu
1 5 10 15

Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn Gly Gly Pro Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser
35

<210> 11
<211> 39
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic construct
<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> xaa at position 1 is L-histidine, D-histidine, or is deleted.

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> Xaa at position 2 is Gly, Ala, or Val;

<220>

x-13991.ST25.txt

<221> MISC_FEATURE
<222> (10)..(10)
<223> Xaa at position 10 is Leu or Val;

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> Xaa at position 12 is Lys or Ser;

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> Xaa at position 13 is Gln or Tyr;

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> Xaa at position 14 is Met or Leu;

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> Xaa at position 16 is Glu or Gln;

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> Xaa at position 17 is Glu or Gln;

<220>
<221> MISC_FEATURE
<222> (19)..(19)
<223> Xaa at position 19 is Val or Ala;

X-13991.ST25.txt

<220>

<221> MISC_FEATURE

<222> (20)..(20)

<223> Xaa at position 20 is Arg or Lys;

<220>

<221> MISC_FEATURE

<222> (21)..(21)

<223> Xaa at position 21 is Leu or Glu;

<220>

<221> MISC_FEATURE

<222> (24)..(24)

<223> Xaa at position 24 is Glu or Ala;

<220>

<221> MISC_FEATURE

<222> (27)..(27)

<223> Xaa at position 27 is Val or Lys;

<220>

<221> MISC_FEATURE

<222> (28)..(28)

<223> Xaa at position 28 is Asn or Lys;

<220>

<221> MISC_FEATURE

<222> (30)..(30)

<223> Xaa at position 30 is Gly or Arg; and

<220>

<221> MISC_FEATURE

<222> (31)..(31)

<223> Xaa at position 31 is Gly, or Pro;

X-13991.ST25.txt

<220>
<221> MISC_FEATURE
<222> (32)..(32)
<223> Xaa at position 32 is Ser, or is absent.

<220>
<221> MISC_FEATURE
<222> (33)..(33)
<223> Xaa at position 33 is Ser, or is absent.

<220>
<221> MISC_FEATURE
<222> (34)..(34)
<223> Xaa at position 34 is Gly, or is absent.

<220>
<221> MISC_FEATURE
<222> (35)..(35)
<223> Xaa at position 35 is Ala, or is absent.

<220>
<221> MISC_FEATURE
<222> (36)..(36)
<223> Xaa at position 36 is Pro, or is absent.

<220>
<221> MISC_FEATURE
<222> (37)..(37)
<223> Xaa at position 37 is Pro, or is absent.

<220>
<221> MISC_FEATURE
<222> (38)..(38)

X-13991.ST25.txt

<223> Xaa at position 38 is Pro, or is absent.

<220>

<221> MISC_FEATURE

<222> (39)..(39)

<223> Xaa at position 39 is Pro, or is absent.

<220>

<221> MISC_FEATURE

<222> (39)..(39)

<223> Xaa at position 39 is Ser, or is absent.

<400> 11

Xaa Xaa Glu Glu Thr Phe Thr Ser Asp Xaa Ser Xaa Xaa Xaa Glu Xaa
1 5 10 15Xaa Ala Xaa Xaa Xaa Phe Ile Xaa Trp Leu Xaa Xaa Gly Xaa Xaa Xaa
20 25 30Xaa Xaa Xaa Xaa Xaa Xaa Xaa
35

<210> 12

<211> 31

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa at position 1 is L-histidine, D-histidine, or is deleted.

<220>

<221> MISC_FEATURE

<222> (2)..(2)

X-13991.ST25.txt

<223> Xaa at position 2 is Ala, Gly, Val, Leu, Ile, Ser, or Thr;

<220>

<221> MISC_FEATURE

<222> (6)..(6)

<223> Xaa at position 6 is Phe, Trp, or Tyr;

<220>

<221> MISC_FEATURE

<222> (10)..(10)

<223> Xaa at position 10 is Val, Trp, Ile, Leu, Phe, or Tyr;

<220>

<221> MISC_FEATURE

<222> (12)..(12)

<223> Xaa at position 12 is Ser, Trp, Tyr, Phe, Lys, Ile, Leu, Val;

<220>

<221> MISC_FEATURE

<222> (13)..(13)

<223> Xaa at position 13 is Tyr, Trp, or Phe;

<220>

<221> MISC_FEATURE

<222> (14)..(14)

<223> Xaa at position 14 is Leu, Phe, Tyr, or Trp;

<220>

<221> MISC_FEATURE

<222> (16)..(16)

<223> Xaa at position 16 is Gly, Glu, Asp, or Lys;

<220>

<221> MISC_FEATURE

x-13991.ST25.txt

<222> (19)..(19)

<223> Xaa at position 19 is Ala, Val, Ile, or Leu;

<220>

<221> MISC_FEATURE

<222> (21)..(21)

<223> Xaa at position 21 is Glu, Ile, or Ala;

<220>

<221> MISC_FEATURE

<222> (24)..(24)

<223> Xaa at position 24 is Ala or Glu;

<220>

<221> MISC_FEATURE

<222> (27)..(27)

<223> Xaa at position 27 is Val or Ile; and

<220>

<221> MOD_RES

<222> (30)..(30)

<223> AMIDATION

<220>

<221> MISC_FEATURE

<222> (31)..(31)

<223> Xaa at position 31 is Gly, His or is absent.

<400> 12

Xaa Xaa Glu Gly Thr Xaa Thr Ser Asp Xaa Ser Xaa Xaa Xaa Glu Xaa
1 5 10 15Gln Ala Xaa Lys Xaa Phe Ile Xaa Trp Leu Xaa Lys Gly Arg Xaa
20 25 30

<210> 13

x-13991.ST25.txt

<211> 616

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 13

His Val Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly
1 5 10 15Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Asp
20 25 30Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu Glu
35 40 45Asn Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln Gln
50 55 60Cys Pro Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr Glu Phe
65 70 75 80Ala Lys Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys Ser
85 90 95Leu His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu Arg
100 105 110Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro Glu
115 120 125Arg Asn Glu Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn Leu Pro
130 135 140Arg Leu Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe His Asp
145 150 155 160Asn Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu Ile Ala Arg Arg
165 170 175His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys Arg Tyr
180 185 190Lys Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala Ala Cys
195 200 205Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser Ser
210 215 220

x-13991.ST25.txt

Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly Glu Arg
225 230 235 240

Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro Lys
245 250 255

Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys Val
260 265 270

His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp Arg
275 280 285

Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser Ser
290 295 300

Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His Cys
305 310 315 320

Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser Leu
325 330 335

Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala Glu
340 345 350

Ala Lys Asp Val Phe Leu Glu Met Phe Leu Tyr Glu Tyr Ala Arg Arg
355 360 365

His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr Tyr
370 375 380

Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu Cys
385 390 395 400

Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro Gln
405 410 415

Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Gln Leu Gly Glu Tyr
420 425 430

Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro Gln
435 440 445

Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys Val
450 455 460

Gly Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys Ala
465 470 475 480

Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His Glu
485 490 495

x-13991.ST25.txt

Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser Leu
500 505 510

Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr Tyr
515 520 525

Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp Ile
530 535 540

Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala Leu
545 550 555 560

Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu Lys
565 570 575

Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys Ala
580 585 590

Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val Ala
595 600 605

Ala Ser Gln Ala Ala Leu Gly Leu
610 615

<210> 14
<211> 631
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic construct
<400> 14

His Val Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Gly
20 25 30

Gly Gly Gly Ser Gly Gly Ser Gly Gly Gly Ser Asp Ala
35 40 45

His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu Glu Asn
50 55 60

Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln Gln Cys
65 70 75 80

x-13991.ST25.txt

Pro Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr Glu Phe Ala
 85 90 95

 Lys Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys Ser Leu
 100 105 110

 His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu Arg Glu
 115 120 125

 Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro Glu Arg
 130 135 140

 Asn Glu Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn Leu Pro Arg
 145 150 155 160

 Leu Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe His Asp Asn
 165 170 175

 Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu Ile Ala Arg Arg His
 180 185 190

 Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys Arg Tyr Lys
 195 200 205

 Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala Ala Cys Leu
 210 215 220

 Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser Ser Ala
 225 230 235 240

 Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly Glu Arg Ala
 245 250 255

 Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro Lys Ala
 260 265 270

 Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys Val His
 275 280 285

 Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp Arg Ala
 290 295 300

 Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys
 305 310 315 320

 Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His Cys Ile
 325 330 335

 Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser Leu Ala
 340 345 350

x-13991.ST25.txt

Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala Glu Ala
 355 360 365

Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg Arg His
 370 375 380 385

Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr Tyr Glu
 385 390 395 400

Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu Cys Tyr
 405 410 415

Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro Gln Asn
 420 425 430

Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Gln Leu Gly Glu Tyr Lys
 435 440 445

Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro Gln Val
 450 455 460

Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys Val Gly
 465 470 475 480

Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys Ala Glu
 485 490 495

Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His Glu Lys
 500 505 510

Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser Leu Val
 515 520 525

Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr Tyr Val
 530 535 540

Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp Ile Cys
 545 550 555 560

Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala Leu Val
 565 570 575

Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu Lys Ala
 580 585 590

Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys Ala Asp
 595 600 605

Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val Ala Ala
 610 615 620

x-13991.ST25.txt

Ser Gln Ala Ala Leu Gly Leu
625 630

<210> 15

<211> 640

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 15

His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser Gly Gly Gly Ser Asp Ala His Lys Ser Glu Val Ala His
50 55 60

Arg Phe Lys Asp Leu Gly Glu Glu Asn Phe Lys Ala Leu Val Leu Ile
65 70 75 80

Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro Phe Glu Asp His Val Lys
85 90 95

Leu Val Asn Glu Val Thr Glu Phe Ala Lys Thr Cys Val Ala Asp Glu
100 105 110

Ser Ala Glu Asn Cys Asp Lys Ser Leu His Thr Leu Phe Gly Asp Lys
115 120 125

Leu Cys Thr Val Ala Thr Leu Arg Glu Thr Tyr Gly Glu Met Ala Asp
130 135 140

Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn Glu Cys Phe Leu Gln His
145 150 155 160

Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu Val Arg Pro Glu Val Asp
165 170 175

Val Met Cys Thr Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys Lys
180 185 190

x-13991.ST25.txt

Tyr Leu Tyr Glu Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro Glu
195 200 205

Leu Leu Phe Phe Ala Lys Arg Tyr Lys Ala Ala Phe Thr Glu Cys Cys
210 215 220

Gln Ala Ala Asp Lys Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu Leu
225 230 235 240

Arg Asp Glu Gly Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys Ala
245 250 255

Ser Leu Gln Lys Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val Ala
260 265 270

Arg Leu Ser Gln Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser Lys
275 280 285

Leu Val Thr Asp Leu Thr Lys Val His Thr Glu Cys Cys His Gly Asp
290 295 300

Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp Leu Ala Lys Tyr Ile Cys
305 310 315 320

Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu Lys Glu Cys Cys Glu Lys
325 330 335

Pro Leu Leu Glu Lys Ser His Cys Ile Ala Glu Val Glu Asn Asp Glu
340 345 350

Met Pro Ala Asp Leu Pro Ser Leu Ala Ala Asp Phe Val Glu Ser Lys
355 360 365

Asp Val Cys Lys Asn Tyr Ala Glu Ala Lys Asp Val Phe Leu Gly Met
370 375 380

Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp Tyr Ser Val Val Leu
385 390 395 400

Leu Leu Arg Leu Ala Lys Thr Tyr Glu Thr Thr Leu Glu Lys Cys Cys
405 410 415

Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys Val Phe Asp Glu Phe
420 425 430

Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn Cys Glu
435 440 445

Leu Phe Glu Gln Leu Gly Glu Tyr Lys Phe Gln Asn Ala Leu Leu Val
450 455 460

x-13991.ST25.txt

Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr Pro Thr Leu Val Glu
 465 470 475 480

Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys Cys Cys Lys His Pro
 485 490 495

Glu Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val Leu
 500 505 510

Asn Gln Leu Cys Val Leu His Glu Lys Thr Pro Val Ser Asp Arg Val
 515 520 525

Thr Lys Cys Cys Thr Glu Ser Leu Val Asn Arg Arg Pro Cys Phe Ser
 530 535 540

Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys Glu Phe Asn Ala Glu
 545 550 555 560

Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu Arg
 565 570 575

Gln Ile Lys Lys Gln Thr Ala Leu Val Glu Leu Val Lys His Lys Pro
 580 585 590

Lys Ala Thr Lys Glu Gln Leu Lys Ala Val Met Asp Asp Phe Ala Ala
 595 600 605

Phe Val Glu Lys Cys Cys Lys Ala Asp Asp Lys Glu Thr Cys Phe Ala
 610 615 620

Glu Glu Gly Lys Lys Leu Val Ala Ala Ser Gln Ala Ala Leu Gly Leu
 625 630 635 640

<210> 16

<211> 624

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 16

His Gly Glu Gly Thr Phe Thr Ser Asp Leu Ser Lys Gln Met Glu Glu
 1 5 10 15

Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn Gly Gly Pro Ser
 20 25 30

X-13991.ST25.txt

Ser Gly Ala Pro Pro Pro Ser Asp Ala His Lys Ser Glu Val Ala His
 35 40 45

Arg Phe Lys Asp Leu Gly Glu Glu Asn Phe Lys Ala Leu Val Leu Ile
 50 55 60

Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro Phe Glu Asp His Val Lys
 65 70 75 80

Leu Val Asn Glu Val Thr Glu Phe Ala Lys Thr Cys Val Ala Asp Glu
 85 90 95

Ser Ala Glu Asn Cys Asp Lys Ser Leu His Thr Leu Phe Gly Asp Lys
 100 105 110

Leu Cys Thr Val Ala Thr Leu Arg Glu Thr Tyr Gly Glu Met Ala Asp
 115 120 125

Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn Glu Cys Phe Leu Gln His
 130 135 140

Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu Val Arg Pro Glu Val Asp
 145 150 155 160

Val Met Cys Thr Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys Lys
 165 170 175

Tyr Leu Tyr Glu Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro Glu
 180 185 190

Leu Leu Phe Phe Ala Lys Arg Tyr Lys Ala Ala Phe Thr Glu Cys Cys
 195 200 205

Gln Ala Ala Asp Lys Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu Leu
 210 215 220

Arg Asp Glu Gly Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys Ala
 225 230 235 240

Ser Leu Gln Lys Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val Ala
 245 250 255

Arg Leu Ser Gln Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser Lys
 260 265 270

Leu Val Thr Asp Leu Thr Lys Val His Thr Glu Cys Cys His Gly Asp
 275 280 285

Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp Leu Ala Lys Tyr Ile Cys
 290 295 300

x-13991.ST25.txt

Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu Lys Glu Cys Cys Glu Lys
 305 310 315 320

Pro Leu Leu Glu Lys Ser His Cys Ile Ala Glu Val Glu Asn Asp Glu
 325 330 335

Met Pro Ala Asp Leu Pro Ser Leu Ala Ala Asp Phe Val Glu Ser Lys
 340 345 350

Asp Val Cys Lys Asn Tyr Ala Glu Ala Lys Asp Val Phe Leu Gly Met
 355 360 365

Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp Tyr Ser Val Val Leu
 370 375 380

Leu Leu Arg Leu Ala Lys Thr Tyr Glu Thr Thr Leu Glu Lys Cys Cys
 385 390 395 400

Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys Val Phe Asp Glu Phe
 405 410 415

Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn Cys Glu
 420 425 430

Leu Phe Glu Gln Leu Gly Glu Tyr Lys Phe Gln Asn Ala Leu Leu Val
 435 440 445

Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr Pro Thr Leu Val Glu
 450 455 460

Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys Cys Cys Lys His Pro
 465 470 475 480

Glu Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val Leu
 485 490 495

Asn Gln Leu Cys Val Leu His Glu Lys Thr Pro Val Ser Asp Arg Val
 500 505 510

Thr Lys Cys Cys Thr Glu Ser Leu Val Asn Arg Arg Pro Cys Phe Ser
 515 520 525

Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys Glu Phe Asn Ala Glu
 530 535 540

Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu Arg
 545 550 555 560

Gln Ile Lys Lys Gln Thr Ala Leu Val Glu Leu Val Lys His Lys Pro
 565 570 575

x-13991.ST25.txt

Lys Ala Thr Lys Glu Gln Leu Lys Ala Val Met Asp Asp Phe Ala Ala
580 585 590

Phe Val Glu Lys Cys Cys Lys Ala Asp Asp Lys Glu Thr Cys Phe Ala
595 600 605

Glu Glu Gly Lys Lys Leu Val Ala Ala Ser Gln Ala Ala Leu Gly Leu
610 615 620

<210> 17

<211> 640

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 17

His Gly Glu Gly Thr Phe Thr Ser Asp Leu Ser Lys Gln Met Glu Glu
1 5 10 15

Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn Gly Gly Pro Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser Gly Gly Gly Ser Asp Ala His Lys Ser Glu Val Ala His
50 55 60

Arg Phe Lys Asp Leu Gly Glu Glu Asn Phe Lys Ala Leu Val Leu Ile
65 70 75 80

Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro Phe Glu Asp His Val Lys
85 90 95

Leu Val Asn Glu Val Thr Glu Phe Ala Lys Thr Cys Val Ala Asp Glu
100 105 110

Ser Ala Glu Asn Cys Asp Lys Ser Leu His Thr Leu Phe Gly Asp Lys
115 120 125

Leu Cys Thr Val Ala Thr Leu Arg Glu Thr Tyr Gly Glu Met Ala Asp
130 135 140

Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn Glu Cys Phe Leu Gln His
145 150 155 160

x-13991.ST25.txt

Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu Val Arg Pro Glu Val Asp
 165 170 175

Val Met Cys Thr Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys Lys
 180 185 190

Tyr Leu Tyr Glu Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro Glu
 195 200 205

Leu Leu Phe Phe Ala Lys Arg Tyr Lys Ala Ala Phe Thr Glu Cys Cys
 210 215 220

Gln Ala Ala Asp Lys Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu Leu
 225 230 235 240

Arg Asp Glu Gly Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys Ala
 245 250 255

Ser Leu Gln Lys Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val Ala
 260 265 270

Arg Leu Ser Gln Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser Lys
 275 280 285

Leu Val Thr Asp Leu Thr Lys Val His Thr Glu Cys Cys His Gly Asp
 290 295 300

Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp Leu Ala Lys Tyr Ile Cys
 305 310 315 320

Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu Lys Glu Cys Cys Glu Lys
 325 330 335

Pro Leu Leu Glu Lys Ser His Cys Ile Ala Glu Val Glu Asn Asp Glu
 340 345 350

Met Pro Ala Asp Leu Pro Ser Leu Ala Ala Asp Phe Val Glu Ser Lys
 355 360 365

Asp Val Cys Lys Asn Tyr Ala Glu Ala Lys Asp Val Phe Leu Gly Met
 370 375 380

Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp Tyr Ser Val Val Leu
 385 390 395 400

Leu Leu Arg Leu Ala Lys Thr Tyr Glu Thr Thr Leu Glu Lys Cys Cys
 405 410 415

Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys Val Phe Asp Glu Phe
 420 425 430

x-13991.ST25.txt

Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn Cys Glu
 435 440 445

Leu Phe Glu Gln Leu Gly Glu Tyr Lys Phe Gln Asn Ala Leu Leu Val
 450 455 460

Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr Pro Thr Leu Val Glu
 465 470 475 480

Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys Cys Cys Lys His Pro
 485 490 495

Glu Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val Leu
 500 505 510

Asn Gln Leu Cys Val Leu His Glu Lys Thr Pro Val Ser Asp Arg Val
 515 520 525

Thr Lys Cys Cys Thr Glu Ser Leu Val Asn Arg Arg Pro Cys Phe Ser
 530 535 540

Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys Glu Phe Asn Ala Glu
 545 550 555 560

Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu Arg
 565 570 575

Gln Ile Lys Lys Gln Thr Ala Leu Val Glu Leu Val Lys His Lys Pro
 580 585 590

Lys Ala Thr Lys Glu Gln Leu Lys Ala Val Met Asp Asp Phe Ala Ala
 595 600 605

Phe Val Glu Lys Cys Cys Lys Ala Asp Asp Lys Glu Thr Cys Phe Ala
 610 615 620

Glu Glu Gly Lys Lys Leu Val Ala Ala Ser Gln Ala Ala Leu Gly Leu
 625 630 635 640

<210> 18

<211> 264

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

x-13991.ST25.txt

<400> 18

His Val Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly
 1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ala
 20 25 30

Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala
 35 40 45

Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro
 50 55 60

Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val
 65 70 75 80

Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val
 85 90 95

Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln
 100 105 110

Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln
 115 120 125

Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala
 130 135 140

Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro
 145 150 155 160

Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr
 165 170 175

Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser
 180 185 190

Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr
 195 200 205

Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr
 210 215 220

Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe
 225 230 235 240

Ser Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys
 245 250 255

Ser Leu Ser Leu Ser Pro Gly Lys
 260

x-13991.ST25.txt

<210> 19
<211> 272
<212> PRT
<213> Artificial Sequence

<220>

<223> synthetic construct
<400> 19

His Val Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Ala Glu Pro Lys Ser Cys Asp Lys Thr
35 40 45

His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser
50 55 60

Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg
65 70 75 80

Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro
85 90 95

Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala
100 105 110

Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val
115 120 125

Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr
130 135 140

Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr
145 150 155 160

Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu
165 170 175

Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys
180 185 190

Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser
195 200 205

x-13991.ST25.txt

Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp
210 215 220

Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser
225 230 235 240

Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala
245 250 255

Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
260 265 270

<210> 20

<211> 264

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 20

His Val Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ala
20 25 30

Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala
35 40 45

Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro
50 55 60

Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val
65 70 75 80

Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val
85 90 95

Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln
100 105 110

Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln
115 120 125

Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala
130 135 140

X-13991.ST25.txt

Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro
145 150 155 160

Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr
165 170 175

Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser
180 185 190

Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr
195 200 205

Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr
210 215 220

Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe
225 230 235 240

Ser Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys
245 250 255

Ser Leu Ser Leu Ser Pro Gly Lys
260

<210> 21

<211> 272

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 21

His Val Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Ala Glu Pro Lys Ser Cys Asp Lys Thr
35 40 45

His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser
50 55 60

Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg
65 70 75 80

x-13991.ST25.txt

Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro
 85 90 95
 Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala
 100 105 110
 Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val
 115 120 125
 Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr
 130 135 140
 Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr
 145 150 155 160
 Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu
 165 170 175
 Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys
 180 185 190
 Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser
 195 200 205
 Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp
 210 215 220
 Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser
 225 230 235 240
 Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala
 245 250 255
 Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
 260 265 270
 <210> 22
 <211> 272
 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> synthetic construct
 <400> 22
 His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
 1 5 10 15

X-13991.ST25.txt

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Ser Ser Gly Ala Ala Glu Pro Lys Ser Cys Asp Lys Thr
35 40 45

His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser
50 55 60

Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg
65 70 75 80

Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro
85 90 95

Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala
100 105 110

Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val
115 120 125

Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr
130 135 140

Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr
145 150 155 160

Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu
165 170 175

Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys
180 185 190

Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser
195 200 205

Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp
210 215 220

Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser
225 230 235 240

Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala
245 250 255

Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
260 265 270

x-13991.ST25.txt

<211> 287

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 23

His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Gly Gly Gly Ser Gly Gly Gly Gly
35 40 45

Ser Gly Gly Gly Ser Ala Glu Pro Lys Ser Cys Asp Lys Thr His
50 55 60

Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser Val
65 70 75 80

Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg Thr
85 90 95

Pro Glu Val Thr Cys Val Val Asp Val Ser His Glu Asp Pro Glu
100 105 110

Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala Lys
115 120 125

Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val Ser
130 135 140

Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr Lys
145 150 155 160

Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr Ile
165 170 175

Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu Pro
180 185 190

Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys Leu
195 200 205

Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser Asn
210 215 220

x-13991.ST25.txt

Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp Ser
225 230 235 240

Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser Arg
245 250 255

Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala Leu
260 265 270

His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
275 280 285

<210> 24

<211> 284

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 24

His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Gly Gly Gly Ser Gly Gly Gly
35 40 45

Ser Gly Gly Gly Ser Ala Glu Ser Lys Tyr Gly Pro Pro Cys Pro
50 55 60

Ser Cys Pro Ala Pro Glu Phe Leu Gly Gly Pro Ser Val Phe Leu Phe
65 70 75 80

Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val
85 90 95

Thr Cys Val Val Val Asp Val Ser Gln Glu Asp Pro Glu Val Gln Phe
100 105 110

Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro
115 120 125

Arg Glu Glu Gln Phe Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr
130 135 140

X-13991.ST25.txt

Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val
145 150 155 160

Ser Asn Lys Gly Leu Pro Ser Ser Ile Glu Lys Thr Ile Ser Lys Ala
165 170 175

Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Gln
180 185 190

Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly
195 200 205

Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro
210 215 220

Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser
225 230 235 240

Phe Phe Leu Tyr Ser Arg Leu Thr Val Asp Lys Ser Arg Trp Gln Glu
245 250 255

Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala Leu His Asn His
260 265 270

Tyr Thr Gln Lys Ser Leu Ser Leu Ser Leu Gly Lys
275 280

<210> 25

<211> 302

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 25

His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Gly Gly Gly Ser Gly Gly Gly Gly
35 40 45

Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
50 55 60

X-13991.ST25.txt

Gly Gly Gly Gly Ser Ala Glu Pro Lys Ser Cys Asp Lys Thr His Thr
65 70 75 80

Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser Val Phe
85 90 95

Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg Thr Pro
100 105 110

Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro Glu Val
115 120 125

Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala Lys Thr
130 135 140

Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val Ser Val
145 150 155 160

Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys
165 170 175

Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser
180 185 190

Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro
195 200 205

Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys Leu Val
210 215 220

Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser Asn Gly
225 230 235 240

Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp Ser Asp
245 250 255

Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser Arg Trp
260 265 270

Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala Leu His
275 280 285

Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
290 295 300

<210> 26

<211> 294

<212> PRT

x-13991.ST25.txt

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 26

His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Gly
20 25 30Gly Gly Gly Ser Gly Gly Ser Gly Gly Gly Ser Gly Gly
35 40 45Gly Gly Ser Gly Gly Ser Gly Gly Gly Ser Ala Glu Pro
50 55 60Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu
65 70 75 80Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp
85 90 95Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp
100 105 110Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly
115 120 125Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn
130 135 140Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp
145 150 155 160Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro
165 170 175Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu
180 185 190Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr Lys Asn
195 200 205Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile
210 215 220Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr
225 230 235 240

x-13991.ST25.txt

Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys
245 250 255

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
260 265 270

Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu
275 280 285

Ser Leu Ser Pro Gly Lys
290

<210> 27

<211> 280

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 27

His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Ser Gly Ala Pro Pro Pro Ser Ala
35 40 45

Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala
50 55 60

Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro
65 70 75 80

Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val
85 90 95

Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val
100 105 110

Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln
115 120 125

Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln
130 135 140

x-13991.ST25.txt

Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala
145 150 155 160

Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro
165 170 175

Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr
180 185 190

Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser
195 200 205

Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr
210 215 220

Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr
225 230 235 240

Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe
245 250 255

Ser Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys
260 265 270

Ser Leu Ser Leu Ser Pro Gly Lys
275 280

<210> 28

<211> 287

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 28

His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Glu
1 5 10 15

Gln Ala Val Lys Glu Phe Ile Ala Trp Leu Ile Lys Gly Arg Gly Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Gly Gly Gly Ser Gly Gly Gly Gly
35 40 45

Ser Gly Gly Gly Ser Ala Glu Pro Lys Ser Cys Asp Lys Thr His
50 55 60

X-13991.ST25.txt

Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser Val
65 70 75 80

Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg Thr
85 90 95

Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro Glu
100 105 110

Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala Lys
115 120 125

Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val Ser
130 135 140

Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr Lys
145 150 155 160

Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr Ile
165 170 175

Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu Pro
180 185 190

Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys Leu
195 200 205

Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser Asn
210 215 220

Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp Ser
225 230 235 240

Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser Arg
245 250 255

Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala Leu
260 265 270

His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
275 280 285

<210> 29

<211> 272

<212> PRT

<213> Artificial Sequence

x-13991.ST25.txt

<220>

<223> synthetic construct

<400> 29

His Gly Glu Gly Thr Phe Thr Ser Asp Leu Ser Lys Gln Met Glu Glu
1 5 10 15Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn Gly Gly Pro Ser
20 25 30Ser Gly Ala Pro Pro Pro Ser Ala Glu Pro Lys Ser Cys Asp Lys Thr
35 40 45His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser
50 55 60Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg
65 70 75 80Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro
85 90 95Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala
100 105 110Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val
115 120 125Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr
130 135 140Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr
145 150 155 160Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu
165 170 175Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys
180 185 190Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser
195 200 205Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp
210 215 220Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser
225 230 235 240Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala
245 250 255

x-13991.ST25.txt

Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
260 265 270

<210> 30

<211> 272

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 30

His Gly Glu Gly Thr Phe Thr Ser Asp Leu Ser Lys Gln Met Glu Glu
1 5 10 15

Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn Gly Gly Pro Ser
20 25 30

Ser Gly Ala Ser Ser Gly Ala Ala Glu Pro Lys Ser Cys Asp Lys Thr
35 40 45

His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser
50 55 60

Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg
65 70 75 80

Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro
85 90 95

Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala
100 105 110

Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val
115 120 125

Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr
130 135 140

Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr
145 150 155 160

Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu
165 170 175

Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys
180 185 190

x-13991.ST25.txt

Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser
195 200 205

Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp
210 215 220

Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser
225 230 235 240

Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala
245 250 255

Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
260 265 270

<210> 31

<211> 287

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 31

His Gly Glu Gly Thr Phe Thr Ser Asp Leu Ser Lys Gln Met Glu Glu
1 5 10 15

Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn Gly Gly Pro Ser
20 25 30

Ser Gly Ala Pro Pro Pro Ser Gly Gly Gly Ser Gly Gly Gly Gly
35 40 45

Ser Gly Gly Gly Ser Ala Glu Pro Lys Ser Cys Asp Lys Thr His
50 55 60

Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser Val
65 70 75 80

Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg Thr
85 90 95

Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro Glu
100 105 110

Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala Lys
115 120 125

x-13991.ST25.txt

Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val Ser
 130 135 140

Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr Lys
 145 150 155 160

Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr Ile
 165 170 175

Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu Pro
 180 185 190

Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val Ser Leu Thr Cys Leu
 195 200 205

Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser Asn
 210 215 220

Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp Ser
 225 230 235 240

Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser Arg
 245 250 255

Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala Leu
 260 265 270

His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
 275 280 285

<210> 32

<211> 232

<212> PRT

<213> Homo sapiens

<400> 32

Ala Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro
 1 5 10 15

Ala Pro Glu Lys Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro
 20 25 30

Lys Asp Thr Lys Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val
 35 40 45

Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val
 50 55 60

x-13991.ST25.txt

Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln
 65 70 75 80

Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln
 85 90 95

Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala
 100 105 110

Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro
 115 120 125

Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr
 130 135 140

Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser
 145 150 155 160

Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr
 165 170 175

Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr
 180 185 190

Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe
 195 200 205

Ser Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys
 210 215 220

Ser Leu Ser Leu Ser Pro Gly Lys
 225 230

<210> 33

<211> 703

<212> DNA

<213> Homo sapiens

<400> 33
 gagcccaaat cttgtgacaa aactcacaca tgcccacgt gcccagcacc tgaactcctg 60
 gggggaccgt cagtcttcct cttcccccua aaacccaagg acaccctcat gatctcccgg 120
 acccctgagg tcacatgcgt ggtggtggac gtgagccacg aagaccctga ggtcaagttc 180
 aactggtagc tggacggcgt ggaggtgcat aatgccaaga caaagccgcg ggaggagcag 240
 tacaacagca cgtaccgtgt ggtcagcgtc ctcaccgtcc tgcaccagga ctggctgaat 300
 ggcaaggagt acaagtgc当地 ggtctccaac aaagccctcc cagccccat cgagaaaacc 360
 58

X-13991.ST25.txt

atctccaaag	ccaaagggca	gccccgagaa	ccacaggtgt	acaccctgcc	cccatcccg	420
gaggagatga	ccaagaacca	ggtcagcctg	acctgcctgg	tcaaaggctt	ctatcccagc	480
gacatcgccg	tggagtggga	gagcaatggg	cagccggaga	acaactacaa	gaccacgcct	540
cccgtgctgg	actccgacgg	ctccttcttc	ctctatagca	agctcaccgt	ggacaagagc	600
aggtggcagc	aggggaacgt	cttctcatgc	tccgtgatgc	atgaggctct	gcacaaccac	660
tacacgcaga	agagcctctc	cctgtctccg	ggtaaatgtat	agt		703

<210> 34

<211> 585

<212> PRT

<213> Homo sapiens

<400> 34

Asp	Ala	His	Lys	Ser	Glu	Val	Ala	His	Arg	Phe	Lys	Asp	Leu	Gly	Glu
1				5				10					15		

Glu	Asn	Phe	Lys	Ala	Leu	Val	Leu	Ile	Ala	Phe	Ala	Gln	Tyr	Leu	Gln
							25					30			

Gln	Cys	Pro	Phe	Glu	Asp	His	Val	Lys	Leu	Val	Asn	Glu	Val	Thr	Glu
							40				45				

Phe	Ala	Lys	Thr	Cys	Val	Ala	Asp	Glu	Ser	Ala	Glu	Asn	Cys	Asp	Lys
						55				60					

Ser	Leu	His	Thr	Leu	Phe	Gly	Asp	Lys	Leu	Cys	Thr	Val	Ala	Thr	Leu
65					70				75			80			

Arg	Glu	Thr	Tyr	Gly	Glu	Met	Ala	Asp	Cys	Cys	Ala	Lys	Gln	Glu	Pro
						85			90			95			

Glu	Arg	Asn	Glu	Cys	Phe	Leu	Gln	His	Lys	Asp	Asp	Asn	Pro	Asn	Leu
							100		105			110			

Pro	Arg	Leu	Val	Arg	Pro	Glu	Val	Asp	Val	Met	Cys	Thr	Ala	Phe	His
						115			120			125			

Asp	Asn	Glu	Glu	Thr	Phe	Leu	Lys	Lys	Tyr	Leu	Tyr	Glu	Ile	Ala	Arg
						130		135			140				

Arg	His	Pro	Tyr	Phe	Tyr	Ala	Pro	Glu	Leu	Leu	Phe	Phe	Ala	Lys	Arg
						145			150			155		160	

Tyr	Lys	Ala	Ala	Phe	Thr	Glu	Cys	Cys	Gln	Ala	Ala	Asp	Lys	Ala	Ala
						165			170			175			

X-13991.ST25.txt

Cys Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser
 180 185 190

Ser Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly Glu
 195 200 205

Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro
 210 215 220

Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys
 225 230 235 240

Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp
 245 250 255

Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser
 260 265 270

Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His
 275 280 285

Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser
 290 295 300

Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala
 305 310 315 320

Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg
 325 330 335

Arg His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr
 340 345 350

Tyr Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu
 355 360 365

Cys Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro
 370 375 380

Gln Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Asn Leu Gly Glu
 385 390 395 400

Tyr Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro
 405 410 415

Gln Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys
 420 425 430

Val Gly Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys
 435 440 445

x-13991.ST25.txt

Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His
450 455 460

Glu Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser
465 470 475 480

Leu Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr
485 490 495

Tyr Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp
500 505 510

Ile Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala
515 520 525

Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu
530 535 540

Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys
545 550 555 560

Ala Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val
565 570 575

Ala Ala Ser Gln Ala Ala Leu Gly Leu
580 585

<210> 35

<211> 1762

<212> DNA

<213> Homo sapiens

```
<400> 35
gatgcgcaca agagtgggt tgctcatcggtttaaagatt tgggagaaga aaatttcaaa
gccttggtgt tgattgcctt tgctcagttat cttcagcagt gtccatttga agatcatgtaa
aatttagtga atgaagtaac tgaatttgca aaaacatgtg ttgctgatga gtcagctgaa
aattgtgaca aatcacttca taccctttt ggagacaaat tatgcacagt tgcaactctt
cgtgaaacct atggtaaat ggctgactgc tgtgcaaaac aagaacctga gagaaatgaa
tgcttcttgc aacacaaaaga tgacaacccca aacctcccc gattggtgag accagaggtt
gatgtgatgt gcactgcttt tcatgacaat gaagagacat ttttggaaaa atacttat
gaaattgcca gaagacatcc ttacttttat gccccggAAC tcctttctt tgctaaaagg
tataaaagctg ctttacaga atgttgccaa gctgctgata aagctgcctg cctgttgcca
aagctcgatg aacttcggga tgaaggaaag gcttcgtctg ccaaacagag actcaagtgt
600
```

x-13991.ST25.txt

gccagtctcc	aaaaatttgg	agaaagagct	ttcaaagcat	gggcagtagc	tcgcctgagc	660
cagagatttc	ccaaagctga	gtttgcagaa	gtttccaagt	tagtgacaga	tcttaccaaa	720
gtccacacgg	aatgctgcc	tggagatctg	cttgaatgtg	ctgatgacag	ggcggacctt	780
gccaagtata	tctgtgaaaa	tcaagattcg	atctccagta	aactgaagga	atgctgtgaa	840
aaacctctgt	tggaaaaatc	ccactgcatt	gccgaagtgg	aaaatgatga	gatgcctgct	900
gacttgcctt	cattagctgc	tgatTTGTT	gaaagtaagg	atgTTTgca	aaactatgct	960
gaggcaaagg	atgtcttcct	gggcatgttt	ttgtatgaat	atgcaagaag	gcatcctgat	1020
tactctgtcg	tgctgctgct	gagacttgcc	aagacatatg	aaaccactct	agagaagtgc	1080
tgtccgctg	cagatcctca	tgaatgctat	gccaaagtgt	tcgatgaatt	taaacctctt	1140
gtggaagagc	ctcagaattt	aatcaaacaa	aattgtgagc	tttttgagca	gcttggagag	1200
tacaaattcc	agaatgcgct	attagttcgt	tacaccaaga	aagtacccca	agtgtcaact	1260
ccaactcttg	tagaggtctc	aagaaaccta	ggaaaagtgg	gcagcaaatg	ttgtaaacat	1320
cctgaagcaa	aaagaatgcc	ctgtgcagaa	gactatctat	ccgtggcct	gaaccagtt	1380
tgtgtgttgc	atgagaaaac	gccagtaagt	gacagagtca	ccaaatgctg	cacagaatcc	1440
ttggtaaca	ggcgaccatg	ctttcagct	ctggaagtgc	atgaaacata	cgttccaaa	1500
gagtttaatg	ctgaaacatt	cacccat	gcagatata	gcacactttc	tgagaaggag	1560
agacaaatca	agaaacaaac	tgcacttgg	gagctcgta	aacacaagcc	caaggcaaca	1620
aaagagcaac	tgaaagctgt	tatggatgat	ttcgagctt	ttgttagagaa	gtgctgcaag	1680
gctgacgata	aggagacctg	cttgccgag	gagggtaaaa	aacttgg	tgcaagtcaa	1740
gctgccttag	gcttataatg	ac				1762